REPORT RESUMES

ED 018 862 80 EA 001 261

THE MEASUREMENT AND EVALUATION OF TEACHING, A CONCEPTUALIZATION OF A PLAN FOR USE IN STATE EDUCATIONAL LEADERSHIP.

BY- DANIEL, K. FRED
FLORIDA ST. DEPT. OF EDUCATION, TALLAHASSEE
REPORT NUMBER ESEA-5 PUB DATE MAY 67
MULTI-STATE TEACHER EDUCATION PROJECT,BALTIMORE,MD
EDRS PRICE MF-$0.75 HC-$5.40 133P.

DESCRIPTORS- *TEACHER EVALUATION, *ELEMENTARY SCHOOL TEACHERS, *SECONDARY SCHOOL TEACHERS, TEACHER EDUCATION, TEACHER CERTIFICATION, ACCREDITATION (INSTITUTIONS), *STATE DEPARTMENTS OF EDUCATION, GOVERNMENT ROLE, TEACHING QUALITY, EFFECTIVE TEACHING, EDUCATIONAL OBJECTIVES, *EVALUATION CRITERIA, RATING SCALES, DATA ANALYSIS, PILOT PROJECTS, INSTITUTIONAL RESEARCH, STATE PROGRAMS, TALLAHASSEE,

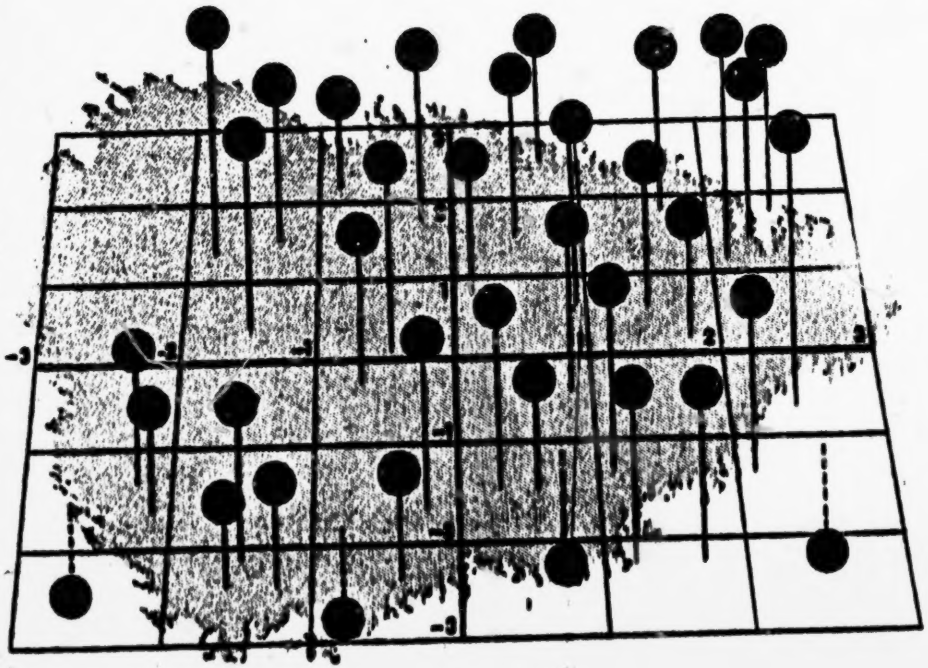
A MAILED QUESTIONNAIRE SURVEY WAS SENT TO THE DIRECTOR OF TEACHER EDUCATION AND/OR CERTIFICATION IN EACH STATE TO DETERMINE THE CONDITIONS UNDER WHICH STATEWIDE PROGRAMS FOR EVALUATING TEACHING CAN YIELD THE MOST USEFUL INFORMATION FOR IMPLEMENTING STATE EDUCATIONAL LEADERSHIP IN THE EDUCATION AND CERTIFICATION OF TEACHERS AND IN THE ACCREDITATION OF SCHOOLS. OF THE COMPLETED QUESTIONNAIRES RECEIVED FROM ALL STATES, RETURNS INDICATED THAT ONLY THREE STATES (FLORIDA, HAWAII, AND PENNSYLVANIA) HAVE ESTABLISHED PROCEDURES FOR EVALUATING TEACHERS THROUGHOUT THE STATE. THE THREE ESTABLISHED PROGRAMS ARE DESCRIBED WITH EXAMPLES INCLUDED OF FORMS USED, BUT NONE OF THE PROCEDURES IS DEEMED ADEQUATE FOR PROVIDING DIRECT PROGRAM GUIDANCE TO FACILITATE STATE EDUCATIONAL LEADERSHIP. SUGGESTIONS ARE GIVEN FOR DETERMINING EVALUATION CRITERIA, THE USE OF SELF-REPORTS, RATING SCALES, AND THE PROCESSES OF OBSERVATION AND EVALUATION. FOR A TEACHER EVALUATION PROGRAM TO BE EFFECTIVE, THE INFORMATION USED MUST BE RELEVANT, INTERPRETABLE, AND RELIABLE. TWO PROPOSALS SUMMARIZE THE STUDY--(1) STATE GOVERNMENTS SHOULD BROADEN THEIR COMMITMENT TO INSTITUTIONAL RESEARCH AS A BASIS FOR DECISIONMAKING IN EDUCATION, AND (2) EDUCATORS IN ELEMENTARY AND SECONDARY SCHOOLS SHOULD BEGIN TO CODIFY THEIR PROFESSIONAL WISDOM. EA 001 269 IS A RELATED DOCUMENT. (JK)

ED018862

The Measurement and Evaluation of Teaching

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.



MULTI-STATE TEACHER EDUCATION PROJECT

A Conceptualization of a Plan for Use in State Educational Leadership

MAY 1967

STATE DEPARTMENT OF EDUCATION
TALLAHASSEE FLORIDA

Floyd T. Christian
SUPERINTENDENT

ACKNOWLEDGEMENTS

This study was carried out in the Florida State Department of Education by K. Fred Daniel. It was supported through the Multi-State Teacher Education Project with funds granted by the U. S. Office of Education under Title V of Public Law 89-10 (Elementary and Secondary Education Act of 1965).

Numerous individuals have assisted in the preparation of this study. Special debts of gratitude are expressed to W. Earl Armstrong, Wm. Cecil Golden, Raymond E. Schultz, and Mary Lou Daniel. The criticism and suggestions provided at critical stages by Edward L. Palmer, Jacob G. Beard, and Francis R. Allen are also appreciated.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

LIST OF ILLUSTRATIONS vi

Chapter

I. THE SIGNIFICANCE AND AMBIGUITY OF INFORMATION ON TEACHING 1

The Significance of Information on Teaching for State Educational Leadership
    Present Use of Information on Teaching for State Educational Leadership
The Ambiguity of Information on Teaching
    Criteria for Evaluating Teaching
    The Status of Knowledge About Effective Teaching
    Outside Factors Affecting Teaching Performance and Outcomes
    The Approach for Reducing Ambiguity Proposed in the Present Study
Conceptual Framework for This Study
    Guidelines
    Measurement, Evaluation and Prediction
    Some Specific Questions
Outline of the Study

II. STATE PRESCRIBED PROCEDURES FOR EVALUATING TEACHING . 19

A Survey of Statewide Teacher Evaluation Programs
The Application of State Prescribed Teacher Evaluation Procedures in Pennsylvania
Statewide Teacher Evaluation in Hawaii
Statewide Teacher Evaluation in Florida
Concluding Statement

Chapter Page

III. APPROACHES TO THE EVALUATION OF TEACHING 47

Selecting Criteria
    Are Some Criteria More Worthwhile than Others?
    Some Clarifications Regarding the Present Study
Collecting Evidence and Comparing it with Criteria
    Self Reports and Self Evaluation
    Rating and Rating Scales
    Separating the Observation Process from the Evaluation Process
    Testing
Concluding Statements

IV. AN APPROACH TO ORGANIZING AND ANALYZING INFORMATION ON TEACHING TO IMPLEMENT STATE EDUCATIONAL LEADERSHIP 74

Providing for Relevance, Interpretability, and Reliability
Collecting and Storing the Data
Analyzing the Data
Initial Steps in Developing a State Program to Collect and Analyze Information on Teaching
    Inducing Local School Districts to Establish Evaluation Programs
    Pilot Projects to Develop Exemplary Programs for Evaluating Teaching
    Disseminating Information on Newly Developed Programs for the Evaluation of Teaching
    Developing the Central Data System

V. RECAPITULATION AND REFLECTIONS 98

Recapitulation
Reflections
    Codifying the Wisdom of Educational Practitioners
    Broadening the State Commitment to Institutional Research
    Some Side Effects
Conclusion

APPENDICES 111

BIBLIOGRAPHY 115

VITA 123

LIST OF TABLES

Table Page

1. Ratings Assigned to 45,035 Florida Teachers for the 1963-64 School Year Using the Florida Teacher Evaluation Form 36

2. Ratings Assigned to 49,977 Florida Teachers for the 1964-65 School Year Using the Florida Teacher Evaluation Form 37

3. Comparisons of Ratings Assigned to the Same Teachers Using the Official Florida Teacher Evaluation Form . . 42

LIST OF ILLUSTRATIONS

Figure | Page

1. "Temporary and Professional Employee's Rating Card" used in Pennsylvania (front and back sides) 23

2. "Report on Principal's Classroom Visit" Used for Probationary Teachers in Hawaii 28

3. "Probationary Teacher Evaluation Report" Used in Hawaii 29

4. "Tenure Teacher Evaluation Report," Part One, Used in Hawaii 30

5. "Tenure Teacher Evaluation Report," Part Two, Used in Hawaii 32

6. Florida Teacher Evaluation Form 36

7. Graphic Representation of System for Classifying Evaluation Procedures Along Three Dimensions 57

8. General Design for the Data Analysis System 86

CHAPTER I

THE SIGNIFICANCE AND AMBIGUITY OF INFORMATION ON TEACHING

The purpose of this study is to delineate the conditions under which statewide programs for measuring and evaluating teaching in elementary and secondary schools can yield optimum information for use by the states in carrying out their educational leadership responsibilities. These responsibilities relate particularly to three state services administered through state departments of education: the education of teachers, the certification of teachers, and the accreditation of schools.

To achieve its purpose, the study describes possible approaches for measuring and evaluating teaching and proposes a plan for developing a state program for collecting and analyzing information on teaching. A review of the significance of such an undertaking, along with problems involved, is presented in the present chapter.

The Significance of Information on Teaching for State Educational Leadership

State educational leadership is defined here as the initiation of new structures or procedures for accomplishing educational objectives of a state or for changing these objectives. This definition is based upon one proposed by Hemphill[1] and used also by Lipham.[2] The latter author contrasts leadership and administration, defining administration as the utilization of existing structures or procedures to accomplish established goals.[3]

State governments have a mandate to carry out both leadership and administrative responsibilities in education, since public education is a state responsibility. The Tenth Amendment of the federal Constitution assigns to the states those powers which are neither reserved to the federal government nor denied to the states. Among these is education. Admittedly, a great deal of discretion has traditionally been exercised by local school corporations; this has been possible only by virtue of authority granted by state governments. Every state possesses a large measure of power in the area of education which it has not yet begun to exercise. Also worthy of note is the fact that the role of the federal government in education has expanded greatly in recent years; such involvement has been possible only under the justification that it promotes the general welfare. Each of the fifty states holds the authority to determine the conditions under which schools shall be established, the qualifications necessary for persons who teach in the schools, and the specific subjects which can and shall be taught in the schools.

---

[1] John K. Hemphill, "Administration as Problem Solving," *Administrative Theory in Education*, ed. Andrew W. Halpin (Chicago: Midwest Administration Center, University of Chicago, 1958), p. 98.

[2] James M. Lipham, "Leadership and Administration," *Behavioral Science and Educational Administration*, The Sixty-third Yearbook of the National Society for the Study of Education, Part II, ed. Daniel E. Griffiths (Chicago: The University of Chicago Press, 1964), p. 122.

[3] *Ibid.*

Each state government exercises leadership by establishing constitutional provisions for an education system and by establishing policies intended to make it possible for each child within the state to secure a high quality education. It is generally acknowledged that the quality of teaching which takes place is an important factor in determining the quality of an educational program. Hence, state leadership has resulted in curriculum guides and consultative services to help improve teaching, plus comprehensive regulatory systems to protect the public from incompetent teaching.

These regulatory systems are comprehensive in that their effects are felt at three different stages: in teacher preparation, at the completion of teacher preparation, and throughout a teacher's tenure. States regulate teacher preparation by requiring that teacher education institutions be granted approval by a state agency.[1] At the completion of pre-service programs teachers in all states must be granted state certificates before they are eligible to teach in public schools.[2] Conditions under which certificates can be renewed or extended are prescribed by the individual states. Accreditation or other regulatory systems are established to prevent teachers from being assigned in areas for which they lack the necessary qualifications and to eliminate other sub-standard teaching situations.

---

[1] This practice is followed in all states except Hawaii, New Mexico, and North Dakota. John R. Mayor and Willis G. Swartz, Accreditation in Teacher Education: Its Influence on Higher Education (Washington, D. C.: National Commission on Accrediting, 1965), p. 23.

[2] W. Earl Armstrong and T. M. Stinnett, A Manual on Certification Requirements for School Personnel in the United States (Washington, D. C.: National Commission on Teacher Education and Professional Standards, National Education Association, 1964).

However, to exercise further leadership through revision of the regulations and improvement of the services, information which is immediately relevant to the services is required. Since the services are directed toward improving teaching, information on the nature of the teaching which is taking place in the state is most relevant. Without such information, leadership must be based upon less significant evidence.

If those persons who exercise control over the services receive no information relative to the effectiveness of their operation, it is probable that no change will be initiated (i.e., no leadership will take place). This idea of organizational inertia is an important postulate in "general systems theory,"[1] a meta-theory or super-theory which has been suggested to provide the framework for concepts in a number of areas and disciplines. The theory provides a series of postulates relative to the functioning of systems. A system is a complex of elements in mutual interaction. These elements may be chemicals, animals, people, groups of people, organizations, nations, planets, or virtually anything else. Griffiths, who is cited above, deals primarily with social organizations, employing the theory as a means of investigating the problem of change in organizations.

---

[1] Daniel E. Griffiths, "Administrative Theory and Change in Organizations," Innovation in Education, ed. Matthew B. Miles (New York: Bureau of Publication, Teachers College, Columbia University, 1964), pp. 425-436.

Central in systems theory is the role of feedback, information received from the environment which can be used in regulating future activities of the system. Through the use of feedback, systems regulate themselves. If the controlling elements of a system receive no relevant feedback, they will have no basis for initiating change.

To revise standards and policies pertaining to the education of school personnel, the certification of school personnel, and the accreditation of schools, information on teaching which is taking place and which has taken place is prime feedback. Measurements and evaluations of teaching could make available qualitative evidence as to the effectiveness of those state services which are established to provide an adequate supply of qualified teachers. These include recruitment activities, in-service education programs, and approval of pre-service education programs. In teacher certification, measurements and evaluations of teaching could provide an indication of the validity of certain aspects of these requirements. In the accreditation of schools, information on the relationship between various school characteristics and quality of teaching could be very useful in developing accreditation standards and procedures.

Present Use of Information on Teaching for State Educational Leadership

As much significance as the potential which information on teaching appears to hold for improving educational leadership decisions, it may seem ironic that virtually no systematic use is presently made of such information by state governments. As the survey reported in Chapter II indicates, only three states have attempted to establish programs for evaluating teaching. Mayor and Swartz report no use of information on teaching performance of graduates in programs for the approval or accreditation of teacher education.[1] No evidence was found of the use of information on teaching in state programs for teacher certification.[2] In the area of school accreditation, Grizzell reports that it is generally acknowledged that present secondary school standards have not been validated, but are largely adaptations of the standards originally adopted by the North Central Association in 1902.[3]

The Ambiguity of Information on Teaching

The principal explanation for the rare use of information on teaching in situations where its potential contributions appear so great must certainly relate to the ambiguity of that information. Virtually any information on teaching which may be gathered is subject to multiple and variant interpretations. Three factors contributing to this situation are (1) differing concepts of what criteria should be used in evaluating teaching, (2) limited understanding of factors which make teaching effective or ineffective, and (3) factors other than teaching which affect the outcomes of teaching.

---

[1] Mayor and Swartz.

[2] Armstrong and Stinnett.

[3] E. D. Grizzell, "Accreditation: Secondary Schools," Encyclopedia of Educational Research, ed. Chester W. Harris (New York: The Macmillan Company, 1960), p. 17.

Criteria for Evaluating Teaching

If information on teaching is to have any utility, there must be available some explicit framework for use in interpreting it. Without this framework, such information is mere description with no extrinsic value. Hence, the first step in interpreting information on teaching consists of selecting relevant criteria with which the information can be evaluated. These criteria must, of course, be pertinent to the purposes which the teaching is expected to fulfill. When different evaluators employ different criteria, their interpretations of the same information on teaching will probably be different. This fact was demonstrated by Anderson.[1] Such differences are one source of ambiguity in information on teaching. In the past, there has been much disagreement as to what criteria of effective teaching should be adopted.

Mitzel[2] proposed three classifications for teaching effectiveness criteria. These can be viewed on a goal-proximity scale: (1) product criteria, (2) process criteria, and (3) presage criteria. The employment of any one of these types of criteria calls for measures of teaching which are different from those required by the other types. Hence, the type of criterion employed dictates the type of information on teaching which is relevant. Product criteria relate directly to the goals toward which teaching is directed: changes in the behavior of pupils. Process criteria comprise types of student and teacher behavior believed to be inherently worthwhile. Presage criteria include traits or experiences of teachers which are conjectured to be relevant to product or process criteria: personality characteristics, knowledge, and academic achievement are illustrations.

---

[1] Harold M. Anderson, "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, Vol. 23 (September, 1954), pp. 41-71.

[2] Harold E. Mitzel, "Teacher Effectiveness," Encyclopedia of Educational Research, ed. Harris, pp. 1482-85.

As convenient as it would be to have a standardized set of criteria of good teaching, it is unlikely that any with wide acceptance will be established in the near future. As indicated above, criteria must be pertinent to the objectives which the teaching is expected to accomplish. When the nature of the teaching goals varies, the criteria for evaluating the teaching may vary also. Such a conclusion was reached by the Committee on the Criteria of Teacher Effectiveness which was established in 1951 by the American Educational Research Association. The committee report states that, ". . . in actual research and service, criteria of teacher effectiveness might differ radically according to culture, level, method, curriculum and the like."[1] Brown thinks that this is not only true, but right. In describing the University of Wisconsin Teacher Competence Project (now the Florida Teacher Competence Project), he referred to the use of variable (and even conflicting) criteria as the most acceptable approach in a democracy.[2] To avoid ambiguity when applying this concept, it is necessary to be aware of the explicit criteria employed by the evaluator in order to interpret his evaluations.

---

[1] H. H. Remmers, et al., "Report of the Committee on the Criteria of Teacher Effectiveness," Review of Educational Research, Vol. 22 (June, 1952), p. 241.

[2] [Bob Burton Brown], "Teacher Competence Project" (Madison, Wisconsin: School of Education, The University of Wisconsin, 1964), p. 4. (Mimeographed.)

In sum, it is necessary to employ definite criteria in interpreting information on teaching. Yet, with many criteria available, it seems doubtful that universal acceptance for any one set can be attained. Hence, in any given teaching situation, it is necessary first to reach agreement as to the criteria which are applicable; only after this is it reasonable to expect agreement in interpreting information on teaching which has taken place in that situation. This same conclusion was reached by Yildirim.[1] He states that, if the disagreement arises in an area of value judgment (i.e., it relates to the criteria which should be applicable), reconciliation is contingent upon the willingness of the parties involved to reconsider their basic premises and to seek consensus. If, on the other hand, it relates to a factual judgment (i.e., it pertains to the specific content of information collected), the disagreement can be resolved by collecting additional information or by some other means of appealing to external measurements or standards.

The Status of Knowledge About Effective Teaching

Information relating to the process or prerequisites for teaching is also subject to multiple interpretations because little is known about what makes the work of an individual teacher effective. Research attempting to elucidate this topic has been extensive, but the findings with practical applicability are meager.

---

[1] Cemal Yildirim, "An Analytic Model for Evaluation of Teacher Competence" (unpublished Ph.D. dissertation, Indiana University, Bloomington, 1963), pp. 69-76.

The volume of studies which have been completed can be appreciated by perusing bibliographies by Domas and Tiedeman[1] (1006 entries) and Morsh and Wilder[2] (362 entries). However, the results of these and other efforts were judged by Remmers to contain little information ". . . that a superintendent of schools can safely employ in hiring a teacher or granting him tenure, that an agency can employ in certifying teachers, or that a teacher education faculty can employ in planning or improving teacher education programs."[3] Turner and Fattu state that, "Seventy years of research on teacher effectiveness have not added much to our systematic knowledge, and it is difficult to see how another seventy can do any more if the same procedures are followed."[4] Other writers expressing dissatisfaction with research results and methodology include Barr and Jones,[1] Mitzel,[2] and Ryans.[3]

---

[1] S. J. Domas and D. V. Tiedeman, "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, Vol. 19 (December, 1950), pp. 101-218.

[2] J. E. Morsh and E. W. Wilder, Identifying the Effective Instructor: A Review of Quantitative Studies, 1900-1952 (Research Bulletin No. AFPTRC-TR-55-44, San Antonio, Texas: United States Air Force Personnel and Training Center, 1954).

[3] H. H. Remmers, et al., "Second Report of the Committee on Criteria of Teacher Effectiveness," Journal of Educational Research, Vol. 46 (May, 1953), p. 657.

[4] Richard L. Turner and Nicholas A. Fattu, "Skill in Teaching, Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research," Bulletin of the School of Education, Indiana University, Vol. 36 (May, 1960), p. iii.

Some writers, however, express optimism in view of current approaches to analyzing relationships between various aspects of teacher behavior and subsequent pupil behavior. Cogan is pessimistic about the past but sees hope for the future.[4] He sees new instruments for data collection brought on by modern technology as a key to improving research on teaching. Mitzel also indicates confidence that recent trends in educational research will ultimately lead to knowledge of patterns of effective teacher behavior.[5] He places his hope in massive research efforts conducted by interdisciplinary teams with ample financial support. Soar feels that a new era of research on teaching may now be underway: ". . . one in which real progress is being made in the specification of dimensions of classroom process that are effective in producing pupil change."[6]

---

[1] Arvil S. Barr and Robert E. Jones, "The Measurement and Prediction of Teacher Efficiency," Review of Educational Research, Vol. 28 (June, 1958), pp. 256-264.

[2] Mitzel, "Teacher Effectiveness," Encyclopedia of Educational Research, ed. Harris, pp. 1481-1486.

[3] David G. Ryans, "Theory Development and the Study of Teacher Behavior," Journal of Educational Psychology, Vol. 47 (December, 1956), pp. 462-475.

[4] Morris L. Cogan, "Research on the Behavior of Teachers: A New Phase," Journal of Teacher Education, Vol. 14 (September, 1963), pp. 238-243.

[5] Harold E. Mitzel, "Can We Measure Good Teaching Objectively?" NEA Journal, Vol. 53 (January, 1964), pp. 35-36.

[6] Robert S. Soar, "Observation Systems and Research on the Teaching-Learning Process," Classroom Interaction Newsletter, Vol. 2 (November, 1966), p. 3.

Variables in Teaching Situations

It is apparent however, that the teaching process is not the only factor which contributes to the behavior of pupils within a given class. If this were the case, the behavior (including achievement) of all pupils in a given class would be the same. There are aspects of every teaching situation over which teachers have no direct control but which modify or otherwise alter the influence which they attempt to assert. These aspects are in many cases attributable to the social and cultural backgrounds of the pupils. Generally, pupils from families which value education and which have provided them with a multitude of enriching educational experiences outside of school tend to attain higher levels of academic achievement.

There are also factors in the physical teaching situation which affect responses of pupils to teaching. The time of day, the weather, or the type of facilities and materials available can make certain types of facilities and materials available can make certain types of teaching behavior more or less suitable and can affect the learning of the pupils.

Of prime significance in a teaching situation is the nature of the objectives toward which the teaching is directed. Objectives vary from school district to school district, from school to school within a district, from classroom to classroom within a school, and even from pupil to pupil within a classroom. As objectives vary, the importance of different types of teaching behavior and different types of pupil behavior will vary also.

Thus, information on teaching which does not account for the context in which the teaching takes place is apt to be ambiguous. Certain types of teaching behavior or teaching outcomes may be desirable in certain situations and undesirable in others.

The Approach for Reducing Ambiguity Proposed in the Present Study

To reduce ambiguity, it is necessary insofar as is possible, to make explicit the implicit factors which contribute to uncertainty in interpreting information on teaching. The approach advocated in the present study is simply to collect such additional information and to consider it when analyzing data on teaching performance or teaching outcomes. The additional data would relate mainly to the criteria which are relevant. Such information would also relate indirectly to situation factors which would presumably be taken into consideration when criteria are established.

Because of the volume of the information to be processed, with numerous clerical problems of sorting and collating likely to ensue, the proposed system was conceived with electronic data processing equipment in mind. The files of information would be maintained on magnetic tapes, discs, or other devices and the analyses would be carried out *via* computer.

Conceptual Framework for This Study

As stated earlier, the purpose of this study is to delineate the conditions under which statewide programs for measuring and evaluating teaching can yield the most useful information for implementing state educational leadership in the education of teachers, certification of teachers, and accreditation of schools. It is not anticipated that such information would be used at the state level to make decisions relative to the quality of individual schools or to the quality of work done by individual teachers. It would be used to evaluate and to improve state policies affecting large numbers of teachers or schools.

The basic aim of this study is to outline a system for assembling at the state level usable information on teaching. Usable information on teaching is defined herein as that information which shows the relationship (congruity or discrepancy) between desired teaching processes and/or products and observed teaching processes and/or products. If good teaching is defined as behaving in a certain manner under certain conditions, information which is usable will show congruity or discrepancies between the criteria and the observed behavior in those situations.

Guidelines

There are a number of established practices which have been accepted as sound by both professional educators and the public which they serve. Those stated below are examples. They will be treated as guidelines and, hence, honored as proposals are developed within the current study.

1. The state should not pre-empt local initiative in determining local educational objectives. It should, however, stimulate local authorities to articulate and justify their objectives.
2. The state should not dictate to professional educators the procedures which they should follow or the specific outcomes which they should seek in their efforts to implement local objectives. It should, however, stimulate members of the profession to articulate and justify their practices.

The present study attempts to outline approaches through which a state can fulfill its responsibilities without usurping the domains of educational policy makers and professional educators within local school districts. In general, it proposes that the state assume leadership in developing systems through which (1) objectives and criteria can be stated more clearly, (2) outcomes and procedures can be described and measured more objectively, and (3) measures of outcomes or procedures can be compared directly with objectives (or criteria) to yield usable information.

Measurement, Evaluation and Prediction

The raw information on teaching which is collected constitutes measurements. A measurement of teaching is defined herein as the end product obtained through analyzing the teaching with a standard classification system (i.e., a scale). This classification system can relate to aspects of either the teaching process or the teaching product.

Measurements, in themselves, are neutral. They merely describe what has happened in terms of a given classification system. In order to determine whether what has happened is good, the measurements must be related to an appropriate set of criteria. This process, the comparing of evidence (measurements) with criteria, constitutes evaluation. This concept has been explicated in some detail by Yildirim.[1] Since it is possible for criteria of effective teaching to vary, it is possible that one set of measurements will represent good teaching when compared with one set of criteria and poor teaching when compared with another set.

---

[1] Yildirim, pp. 49-76.

Prediction is an attempt to estimate in advance what given measurements will be. Programs for the certification of teachers constitute an application of prediction. These programs are operated on the assumption that the criteria which they employ are effective in discriminating between those who, if allowed to teach, would tend to receive satisfactory ratings and those who would not. Certificates are then issued only to those in the first group.

The present study deals first with measurements by reviewing and discussing measures of teaching which have been or could be applied. It then treats evaluation by discussing ways in which these measures might be related to evaluative criteria (or ways in which evaluative criteria might be related to these measures). Such a process would thus provide a basis for testing the validity of predictors now employed. This study does not, however, set out to validate or invalidate any specific standards or programs. The purpose herein is to set forth a framework or general system which can be employed to facilitate the use of information on teaching for implementing state educational leadership.

Some Specific Questions

The proposed system is designed to include information which is relevant and accessible for answering several important questions relative to teacher education, teacher certification, and school accreditation. This information would also be usable for a number of other purposes. The questions in the following list are illustrative:

(1) What criteria do individual colleges think are relevant for evaluating the teaching of graduates of their programs? Are these the same for all graduates? For all graduates in given teaching fields?

(2) How does the teaching done by graduates of different types of teacher education programs measure up against the evaluation criteria considered appropriate by the teacher education institution? By the local school districts?

(3) What criteria do individual school districts or schools judge to be relevant for evaluating teaching in their classrooms? Are these the same for all or most teachers within a school district? Within a school? For all or most teachers of certain grades or subjects?

(4) How does the teaching done by teachers in individual schools measure up against the evaluative criteria which are set forth as appropriate by the local school district?

(5) How do the criteria specified for different segments of the teaching population (e.g., beginning teachers, music teachers, sixth grade teachers, etc.) compare?

(6) How do evaluation criteria selected by school districts and evaluation criteria selected by colleges compare?

(7) Is there more than one distinctive pattern of expectations for teachers in a given field?

(8) What are the inservice education needs of various segments of the teaching population?

(9) What are the relationships between teacher status characteristics[1] (viz., the things which are considered in recommending a baccalaureate candidate for a teaching certificate or in issuing a teacher certificate) and teaching (evaluated in terms of criteria established by local school districts)?

(10) What relationships exist between school status characteristics (viz., the things which are given consideration in recommending a school for accreditation) and the teaching and learning which takes place in those schools?

---

[1] Teacher status characteristics which are considered to be necessary antecedents to effective teaching constitute presage criteria.

Outline of the Study

Chapter II reports a questionnaire survey of statewide programs for evaluating teaching. It also contains descriptions of the programs in three states (Hawaii, Pennsylvania and Florida) in which the state prescribes procedures for evaluating teachers. In Chapter III, general approaches for evaluating teaching are reviewed. Chapter IV proposes a system for obtaining information on teaching which could be used to implement state educational leadership. Summarizing and concluding remarks are presented in Chapter V.

CHAPTER II

STATE PRESCRIBED PROCEDURES FOR EVALUATING TEACHING

Comprehensive information on the quality of teaching within a state can contribute significantly to the process of state educational leadership. However, as pointed out in Chapter I, any effort to evaluate teaching is, from the start, plagued with complex problems. There are many different concepts regarding the nature of acceptable criteria for evaluating teaching. Possibly because of this, research which should be expected to provide guidance has made few significant contributions to current practices in evaluating teaching. In addition, there are numerous factors within the context of the teaching situation which affect both the processes and products of teaching.

Educational literature contains virtually no mention of statewide efforts to evaluate teaching. Only two references containing information on this subject were found in the literature: one by Rasley and one by Kelley. Rasley,[1] through a survey of teacher evaluation programs, found that Hawaii, Florida, and Pennsylvania require that teachers be evaluated. An article by Kelley[1] describes the statewide program for evaluating teachers in Florida.

---

[1] Charles W. Rasley, "TER, A Pillar of Strength: A Current View of Teacher Effectiveness Rating as Used in the Public Schools" (A Survey Research Study, Division of Education in the Graduate School, Southern Illinois University, Edwardsville, 1966), p. 39.

Since it was not clear from Rasley's study whether states other than the three cited were polled in his survey, a new questionnaire survey was carried out. The results (identical with Rasley's from the standpoint of states included) are reported below.

A Survey of Statewide Teacher Evaluation Programs

Purpose.--The purpose of the survey was to find out which states have within the past ten years carried out a program of statewide evaluation of teaching and to ascertain the purposes which those evaluations were intended to fulfill. The evaluation with which this survey was concerned is defined as an attempt by a state government or its agent to secure a composite assessment of the quality of the teaching which is taking place within the state.

The questionnaire.--The questionnaire was a one-page instrument containing seven questions. A copy is included as an appendix. The first three questions asked (a) whether a statewide evaluation of teaching has been conducted within the last ten years, (b) whether such evaluations are conducted annually, and (c) when the three most recent evaluations were conducted. The final four questions asked about the purposes for conducting each of the three most recent evaluations. The purposes specifically mentioned were (a) to recognize and promote meritorious teaching, (b) to recognize and eliminate incompetent teaching, and (c) to accumulate research data.

---

[1] J. T. Kelley, "Teacher Evaluation," Florida School Bulletin, Vol. 24 (December, 1961), pp. 23-28.

Population.--The population surveyed included the state director of teacher education and/or certification in each state. The names were obtained from the 1965-66 Roster of State Directors of Teacher Education and Certification.[1] Only one questionnaire was sent to each state; the recipient was the one whose title implied that he would be most likely to have responsibility for a statewide evaluation program in his state, if one existed.

Procedures and returns.--A cover letter was written and signed by W. Cecil Golden, Director, Division of Teacher Education, Certification and Accreditation, Florida State Department of Education. A copy of this letter is included as an appendix. A stamped return envelope, addressed to Golden, was enclosed with the questionnaire and the letter. Individually typed follow-up letters, signed by Golden, were sent in September and again in November to all persons in the population who had not yet returned their questionnaires at those times. Enclosed with each follow-up letter was an additional copy of the questionnaire and a stamped return envelope. Eventually, completed questionnaires were received from all states.

Follow-up of the questionnaire survey.--Those persons who indicated that their states did carry out statewide teacher evaluation programs were contacted by mail, by telephone, or in person to obtain additional information as to their procedures and purposes for these programs. Through this approach, it was found that some of the states which responded affirmatively to the first question did not, in fact, have an evaluation program which would fall within the limitations of the definition employed in the present study. The approach in these states was usually to conduct surveys to determine either the educational levels attained by their teachers or the extent to which certification requirements were being met.

---

[1] Roster of State Directors of Teacher Education and Certification (Washington, D. C.: The National Commission on Teacher Education and Professional Standards, 1965).

Results of the questionnaire survey.--Questionnaire returns indicated that only three states--Pennsylvania, Hawaii, and Florida--prescribe procedures for evaluating teachers to be used throughout the state. The programs in these states are described in the subsequent sections of this chapter. This information was obtained through contacts following the questionnaire survey.

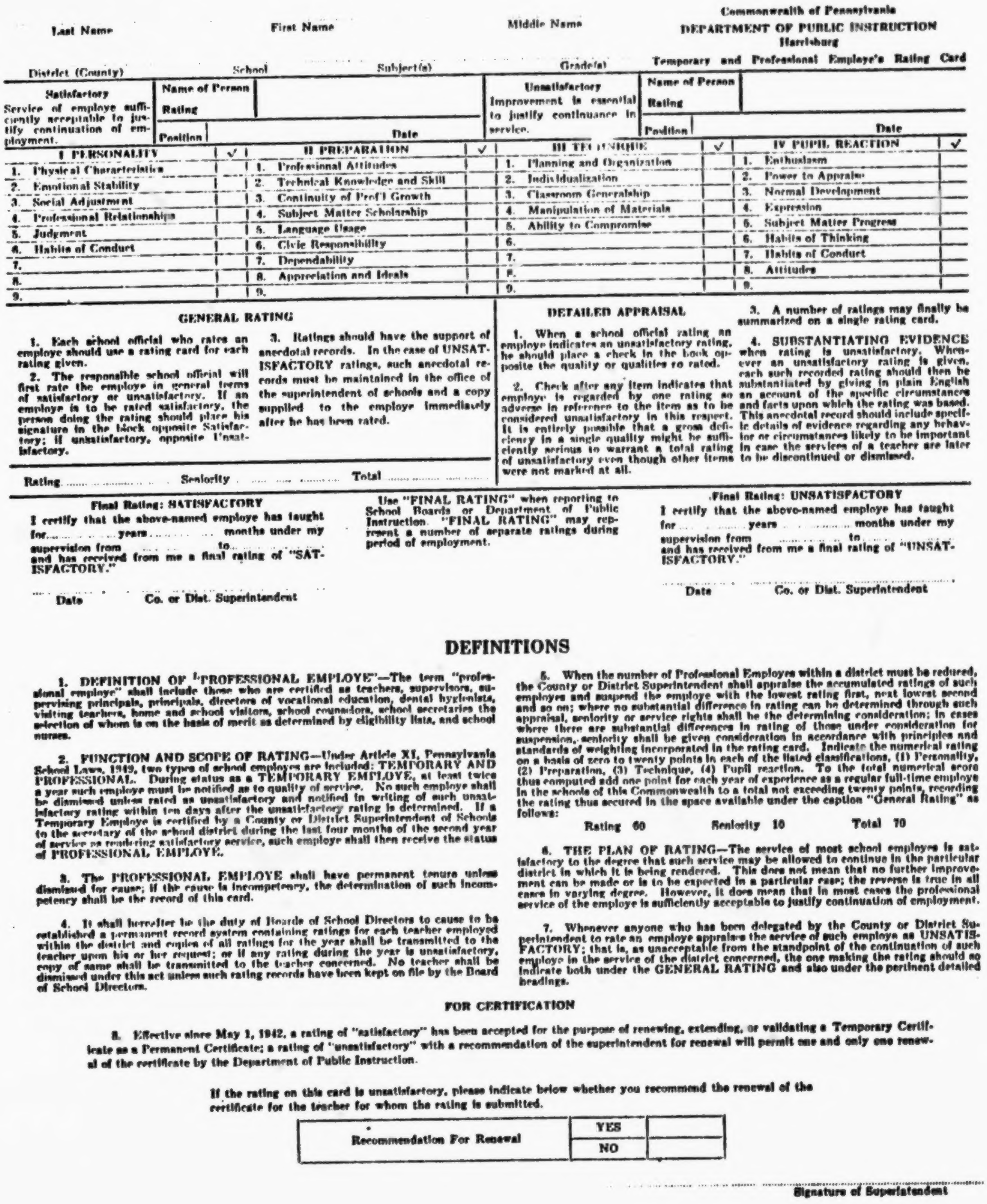
The Application of State Prescribed Teacher Evaluation Procedures in Pennsylvania[1]

In Pennsylvania, a state prescribed rating system is in effect, employing the form depicted in Figure 1. The rating of teachers is a statutory requirement[2] designed to eliminate incompetent teachers. The rating form is prepared by the Department of Public Instruction. The responsibility for rating teachers rests with the County or District Superintendent, but he may delegate this authority to a lower administrative or supervisory officer. Unsatisfactory ratings assigned by

---

[1] The description in this section is based upon information furnished by Harris W. Reynolds, Education Evaluation Advisor, Pennsylvania Department of Public Instruction.

[2] School Laws of Pennsylvania, Article XI, Section 1123, as cited by Reynolds in a personal letter to W. Cecil Golden, August 19, 1966.



Last Name First Name Middle Name

Commonwealth of Pennsylvania
DEPARTMENT OF PUBLIC INSTRUCTION
Harrisburg

District (County) School Subject(s) Grade(s)

Temporary and Professional Employe's Rating Card

Satisfactory Service of employe sufficiently acceptable to justify continuation of employment.	Name of Person Rating		**Unsatisfactory** Improvement is essential to justify continuance in service.	Name of Person Rating	
	Position	Date		Position	Date

I PERSONALITY	✓	II PREPARATION	✓	III TECHNIQUE	✓	IV PUPIL REACTION	✓
1. Physical Characteristics		1. Professional Attitudes		1. Planning and Organization		1. Enthusiasm	
2. Emotional Stability		2. Technical Knowledge and Skill		2. Individualization		2. Power to Appraise	
3. Social Adjustment		3. Continuity of Prof'l Growth		3. Classroom Generalship		3. Normal Development	
4. Professional Relationships		4. Subject Matter Scholarship		4. Manipulation of Materials		4. Expression	
5. Judgment		5. Language Usage		5. Ability to Compromise		5. Subject Matter Progress	
6. Habits of Conduct		6. Civic Responsibility		6.		6. Habits of Thinking	
7.		7. Dependability		7.		7. Habits of Conduct	
8.		8. Appreciation and Ideals		8.		8. Attitudes	
9.		9.		9.		9.	

GENERAL RATING

1. Each school official who rates an employe should use a rating card for each rating given.

2. The responsible school official will first rate the employe in general terms of satisfactory or unsatisfactory. If an employe is to be rated satisfactory, the person doing the rating should place his signature in the block opposite Satisfactory; if unsatisfactory, opposite Unsatisfactory.

3. Ratings should have the support of anecdotal records. In the case of UNSATISFACTORY ratings, such anecdotal records must be maintained in the office of the superintendent of schools and a copy supplied to the employe immediately after he has been rated.

Rating .... Seniority .... Total ....

DETAILED APPRAISAL

1. When a school official rating an employe indicates an unsatisfactory rating, he should place a check in the book opposite the quality or qualities so rated.

2. Check after any item indicates that employe is regarded by one rating so adverse in reference to the item as to be considered unsatisfactory in this respect. It is entirely possible that a gross deficiency in a single quality might be sufficiently serious to warrant a total rating of unsatisfactory even though other items were not marked at all.

3. A number of ratings may finally be summarized on a single rating card.

4. SUBSTANTIATING EVIDENCE when rating is unsatisfactory. Whenever an unsatisfactory rating is given, each such recorded rating should then be substantiated by giving in plain English an account of the specific circumstances and facts upon which the rating was based. This anecdotal record should include specific details of evidence regarding any behavior or circumstances likely to be important in case the services of a teacher are later to be discontinued or dismissed.

Final Rating: SATISFACTORY

I certify that the above-named employe has taught for............ years............ months under my supervision from .... to............ and has received from me a final rating of "SATISFACTORY."

....
Date Co. or Dist. Superintendent

Use "FINAL RATING" when reporting to School Boards or Department of Public Instruction. "FINAL RATING" may represent a number of separate ratings during period of employment.

Final Rating: UNSATISFACTORY

I certify that the above-named employe has taught for............ years............ months under my supervision from .... to............ and has received from me a final rating of "UNSATISFACTORY."

....
Date Co. or Dist. Superintendent

DEFINITIONS

1. DEFINITION OF "PROFESSIONAL EMPLOYE"—The term "professional employe" shall include those who are certified as teachers, supervisors, supervising principals, principals, directors of vocational education, dental hygienists, visiting teachers, home and school visitors, school counselors, school secretaries the selection of whom is on the basis of merit as determined by eligibility lists, and school nurses.

2. FUNCTION AND SCOPE OF RATING—Under Article XI, Pennsylvania School Laws, 1949, two types of school employes are included: TEMPORARY AND PROFESSIONAL. During status as a TEMPORARY EMPLOYE, at least twice a year such employe must be notified as to quality of service. No such employe shall be dismissed unless rated as unsatisfactory and notified in writing of such unsatisfactory rating within ten days after the unsatisfactory rating is determined. If a Temporary Employe is certified by a County or District Superintendent of Schools to the secretary of the school district during the last four months of the second year of service as rendering satisfactory service, such employe shall then receive the status of PROFESSIONAL EMPLOYE.

3. The PROFESSIONAL EMPLOYE shall have permanent tenure unless dismissed for cause; if the cause is incompetency, the determination of such incompetency shall be the record of this card.

4. It shall hereafter be the duty of Boards of School Directors to cause to be established a permanent record system containing ratings for each teacher employed within the district and copies of all ratings for the year shall be transmitted to the teacher upon his or her request; or if any rating during the year is unsatisfactory, copy of same shall be transmitted to the teacher concerned. No teacher shall be dismissed under this act unless such rating records have been kept on file by the Board of School Directors.

5. When the number of Professional Employes within a district must be reduced, the County or District Superintendent shall appraise the accumulated ratings of such employes and suspend the employe with the lowest rating first, next lowest second and so on; where no substantial difference in rating can be determined through such appraisal, seniority or service rights shall be the determining consideration; in cases where there are substantial differences in rating of those under consideration for suspension, seniority shall be given consideration in accordance with principles and standards of weighting incorporated in the rating card. Indicate the numerical rating on a basis of zero to twenty points in each of the listed classifications, (1) Personality, (2) Preparation, (3) Technique, (4) Pupil reaction. To the total numerical score thus computed add one point for each year of experience as a regular full-time employe in the schools of this Commonwealth to a total not exceeding twenty points, recording the rating thus secured in the space available under the caption "General Rating" as follows:

Rating 60 Seniority 10 Total 70

6. THE PLAN OF RATING—The service of most school employes is satisfactory to the degree that such service may be allowed to continue in the particular district in which it is being rendered. This does not mean that no further improvement can be made or is to be expected in a particular case; the reverse is true in all cases in varying degree. However, it does mean that in most cases the professional service of the employe is sufficiently acceptable to justify continuation of employment.

7. Whenever anyone who has been delegated by the County or District Superintendent to rate an employe appraises the service of such employe as UNSATISFACTORY; that is, as unacceptable from the standpoint of the continuation of such employe in the service of the district concerned, the one making the rating should so indicate both under the GENERAL RATING and also under the pertinent detailed headings.

FOR CERTIFICATION

8. Effective since May 1, 1942, a rating of "satisfactory" has been accepted for the purpose of renewing, extending, or validating a Temporary Certificate as a Permanent Certificate; a rating of "unsatisfactory" with a recommendation of the superintendent for renewal will permit one and only one renewal of the certificate by the Department of Public Instruction.

If the rating on this card is unsatisfactory, please indicate below whether you recommend the renewal of the certificate for the teacher for whom the rating is submitted.

Recommendation For Renewal	YES	
	NO	

....
Signature of Superintendent

Fig. 1--"Temporary and Professional Employes' Rating Card" Used in Pennsylvania (front and back sides).

such officers are official, however, only when signed by a commissioned officer of the Commonwealth of Pennsylvania (viz., a County or District Superintendent, Assistant Superintendent or Associate Superintendent). Permanent records of all ratings for each teacher are maintained in the teacher's school district. All ratings of a teacher must be transmitted to him at his request. An unsatisfactory rating will be transmitted automatically.

Both beginning and tenure teachers are evaluated. A teacher who enters the teaching profession in Pennsylvania must serve a two-year, or sometimes three-year, probationary period. During this time he is termed a "temporary professional employe" and is entitled to neither contract nor tenure status. "Temporary professional employes" are evaluated twice annually. At the termination of two years of service with "satisfactory" ratings the teacher is eligible to become a "professional employe" with both contract and tenure status. "Professional employes" are evaluated once each year. A "professional employe" may be dismissed only for cause; if the cause is incompetency, the rating form provides the necessary record.

The ratings assigned teachers are also given consideration in the renewal of teaching certificates. Only those teachers with ratings of "satisfactory" are eligible to have their certificates renewed.

Since the rating form is intended to identify and eliminate incompetent teachers and since a rating of "unsatisfactory" on this form will, in fact, prevent a teacher from achieving tenure status, considerable care is exercised in prescribing procedures which must be followed when assigning an "unsatisfactory" rating: The rating form must, of course, be completed and signed properly. In addition, substantiating evidence must be provided, stating "in plain English" the specific conditions or instances upon which the rating is based. On the other hand, if a teacher is rated "satisfactory," it is not necessary to provide, either on the form or elsewhere, information which indicates the teacher's relative strengths or weaknesses.

As for the value of the Pennsylvania program, Department of Education officials from that state feel that it is effective in providing a means of eliminating teachers who have glaring weaknesses. These are teachers who are quite obviously inadequate. It is not designed, however, to provide information for use in making educational decisions (other than decisions as to whether, in specific cases, teaching certificates should be issued).

Statewide Teacher Evaluation in Hawaii[1]

The plan of school organization in Hawaii is unique in that the state, itself, operates all schools, rather than delegating this authority to local districts. Thus, Hawaii's approach to teacher evaluation is in many ways similar to one which might be taken in a large local school district.

---

[1] The description in this section is based upon information furnished by Minoru Ezaki, Staff Specialist, Personnel Development, and Harry Chang, Staff Specialist, Recruitment and Placement, Hawaii State Department of Education through personal correspondence and consultation, 1966 and 1967.

In Hawaii, evaluation of teachers is required by law.[1] The stated purpose of such evaluation is to improve instruction in the schools of the State. To fulfill this purpose, separate forms and procedures for the evaluation of probationary teachers and the evaluation of tenure teachers have been developed. In both cases, the primary responsibility for evaluating the teacher rests with the principal. Both the principal and the teacher who is evaluated must sign the form. (The signature of the teacher is not meant to imply that he concurs with the evaluation but merely that he is aware of it.) Following up on recommendations made to the teacher is primarily the duty of the principal. However, in cases where the rating is unsatisfactory, the district office also takes action.

Reports of the evaluations are placed on file in the State Department of Education where they are deposited in folders of individual teachers and become part of the teachers' permanent personnel records. Copies of the completed forms are also kept in the files of the district superintendent and of the principal who made the evaluation. An additional copy is given to the teacher who is evaluated.

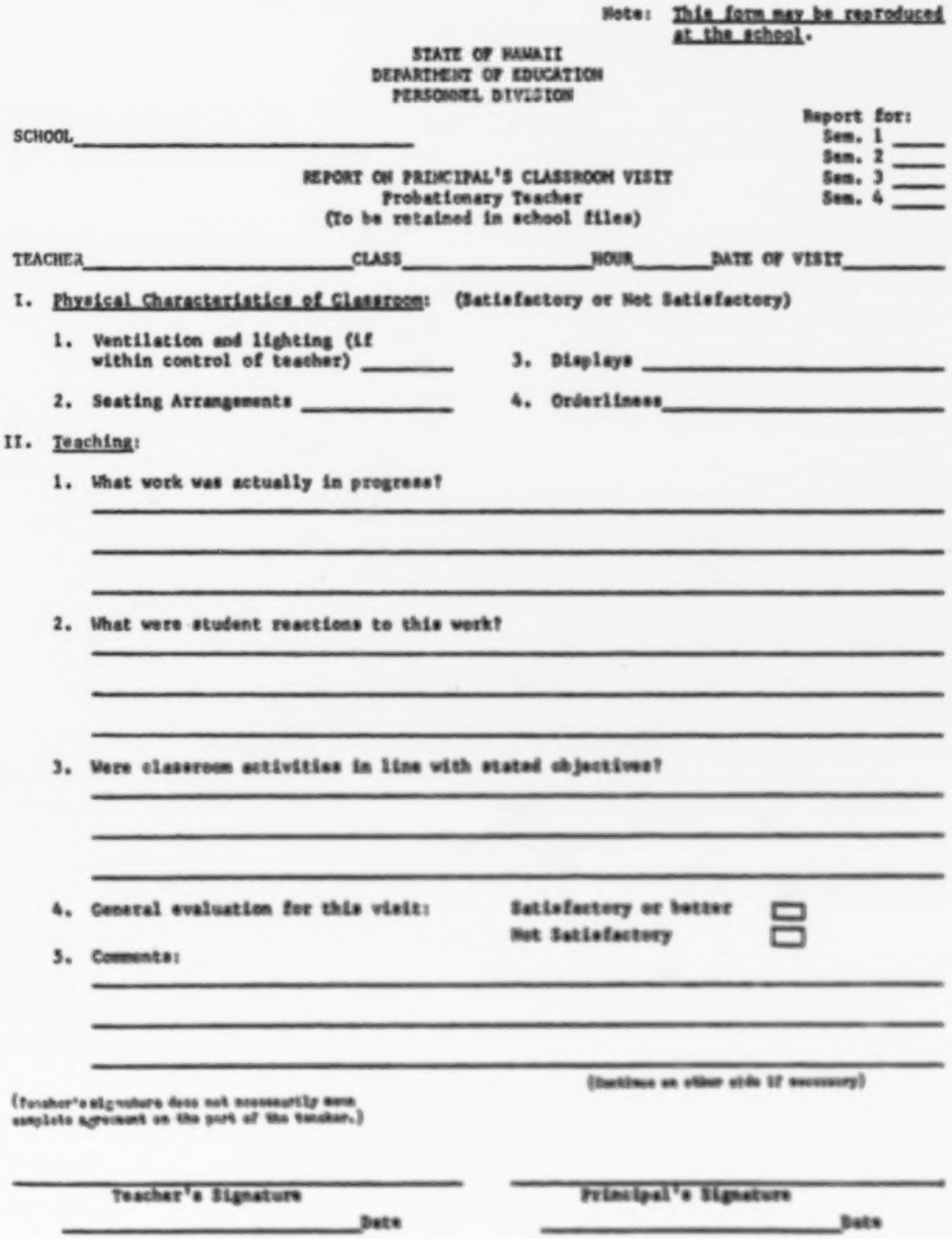
Every teacher beginning service in Hawaii must serve a probationary period of four consecutive semesters. This period may be extended to five years. During each semester, the principal pays several visits (four or five are suggested) to classes of each probationary teacher. After each visit, the principal completes the form,

---

[1] Revised Laws of Hawaii, Section 38-38, as amended (1965), cited in "Personnel Memo No. 318 (1965-66)" (Honolulu, Hawaii: State of Hawaii, Department of Education, October 22, 1965, mimeographed).

"Report on Principal's Classroom Visit" (Figure 2), and reviews it with the teacher. Three copies of the form are filled out: one copy to be retained by the principal, one to be given to the teacher, and one to be forwarded to the district office. Near the close of the semester, the principal prepares a report (Figure 3) summarizing his prior observations. This report is also reviewed and signed by both the principal and the teacher. It is prepared in quadruplicate, with the original being submitted to the Office of Personnel Services, State Department of Education and with the remaining copies being distributed as were the classroom visit reports. When the summary report following the fourth semester of teaching is submitted to the Department, proceedings to grant tenure are initiated. If all four reports are satisfactory, tenure will be granted automatically. If they are not, the case will be investigated. This will involve consultation between the principal, the district superintendent, the State Department of Education personnel officers, and, if necessary, with the teacher also.

Tenure teachers are to be evaluated at least once per year. The evaluation report has two parts: one to be filled out by the teacher which deals with conditions affecting teacher performance (Figure 4) and one to be filled out by the principal which deals with the teacher's professional qualities (Figure 5). Both forms are prepared in quadruplicate and signed by the teacher and the principal, with copies being distributed to the State Office of Personnel Services, the teacher, the principal, and the district office.



Note: This form may be reproduced at the school.

STATE OF HAWAII
DEPARTMENT OF EDUCATION
PERSONNEL DIVISION

SCHOOL ______________________

Report for:
Sem. 1 ____
Sem. 2 ____
Sem. 3 ____
Sem. 4 ____

REPORT ON PRINCIPAL'S CLASSROOM VISIT
Probationary Teacher
(To be retained in school files)

TEACHER ______________ CLASS __________ HOUR ______ DATE OF VISIT ______

I. Physical Characteristics of Classroom: (Satisfactory or Not Satisfactory)

1. Ventilation and lighting (if within control of teacher) ________
2. Seating Arrangements ________
3. Displays ________
4. Orderliness ________

II. Teaching:

1. What work was actually in progress?

2. What were student reactions to this work?

3. Were classroom activities in line with stated objectives?

4. General evaluation for this visit: Satisfactory or better ☐ Not Satisfactory ☐

5. Comments:

(Continue on other side if necessary)

(Teacher's signature does not necessarily mean complete agreement on the part of the teacher.)

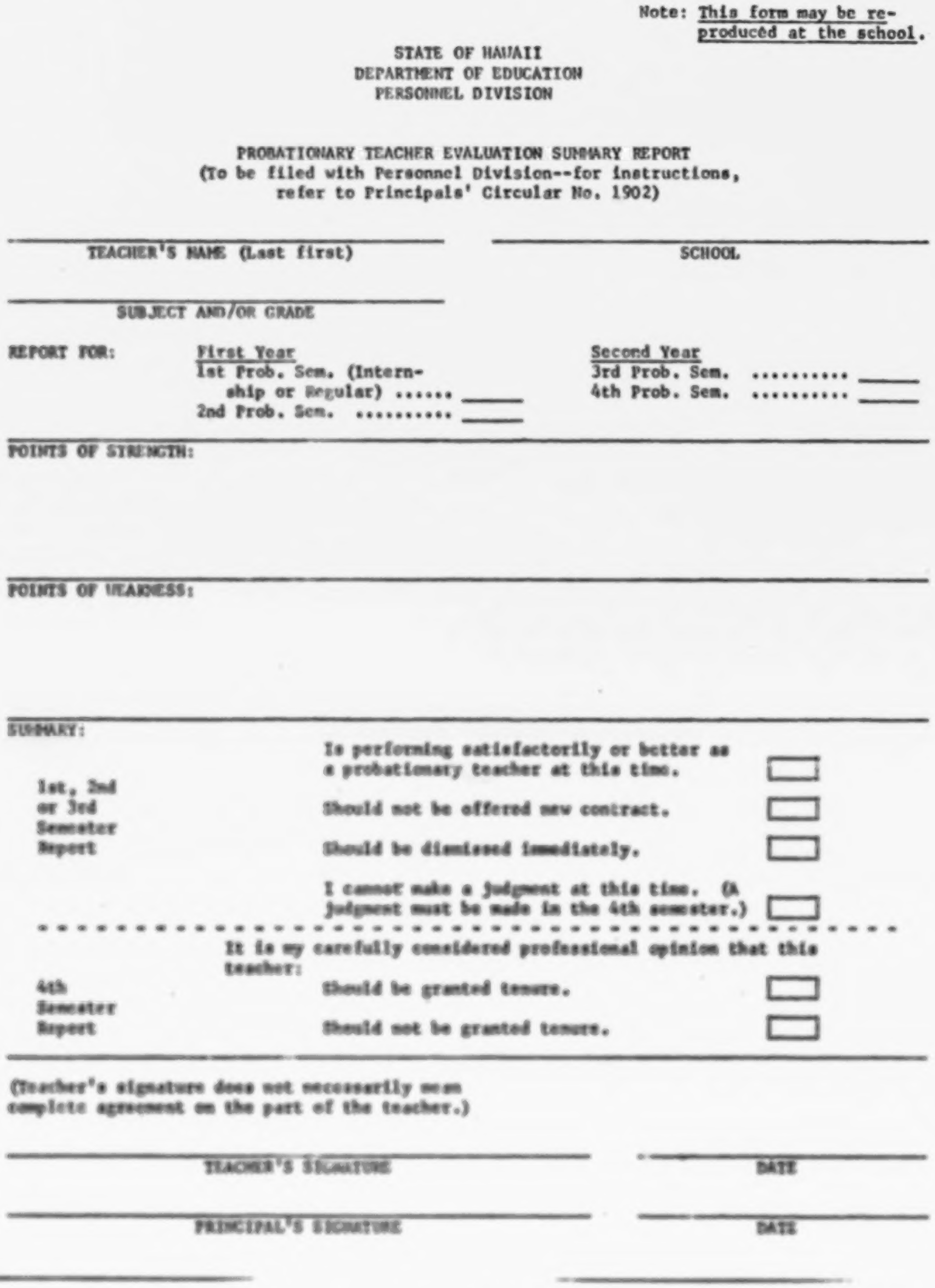
______________________ Teacher's Signature

______________________ Date

______________________ Principal's Signature

______________________ Date

Fig. 2--"Report on Principal's Classroom Visit" Used for Probationary Teachers in Hawaii.



Note: This form may be reproduced at the school.

STATE OF HAWAII
DEPARTMENT OF EDUCATION
PERSONNEL DIVISION

PROBATIONARY TEACHER EVALUATION SUMMARY REPORT
(To be filed with Personnel Division--for instructions, refer to Principals' Circular No. 1902)

TEACHER'S NAME (Last first) ______ SCHOOL ______

SUBJECT AND/OR GRADE ______

REPORT FOR:

First Year
1st Prob. Sem. (Internship or Regular) ...... ______
2nd Prob. Sem. .... ______

Second Year
3rd Prob. Sem. .... ______
4th Prob. Sem. .... ______

POINTS OF STRENGTH:

POINTS OF WEAKNESS:

SUMMARY:

1st, 2nd or 3rd Semester Report

Is performing satisfactorily or better as a probationary teacher at this time. ☐

Should not be offered new contract. ☐

Should be dismissed immediately. ☐

I cannot make a judgment at this time. (A judgment must be made in the 4th semester.) ☐

- -

4th Semester Report

It is my carefully considered professional opinion that this teacher:

Should be granted tenure. ☐

Should not be granted tenure. ☐

(Teacher's signature does not necessarily mean complete agreement on the part of the teacher.)

TEACHER'S SIGNATURE ______ DATE ______

PRINCIPAL'S SIGNATURE ______ DATE ______

Fig. 3--"Probationary Teacher Evaluation Report" Used in Hawaii.

STATE OF HAWAII
DEPARTMENT OF EDUCATION
HONOLULU, HAWAII

8/19/65

TENURE TEACHER EVALUATION REPORT

Year 19___ - 19___

TEACHER______________________________
Last, First Mr., Mrs., Miss

Number of years with Dept. of Educ.__________

SCHOOL______________________________

Grade or Subject__________

PART ONE - CONDITIONS OF WORK AFFECTING TEACHER PERFORMANCE
(To be Filled in by Teacher and Discussed with Principal)
This is an opportunity for the teacher to describe conditions that affect his teaching. You are invited to make comments that apply to you. Return to your Principal before the end of the 1st quarter in November (4 copies).

(Check or Fill in Information)

I. CLASSROOM ENVIRONMENT

TEACHER'S COMMENTS AND SUGGESTIONS
(Comment on Items that Apply to You)

A. Students

1. Ability: Fast____; Average____; Slow____; Mixed Group____.
2. Stability of Enrollment: Stable____; Transient____.
3. Special Problems: Gifted____; M.R.____; Physically Handicapped____; Emotionally Disturbed____; Other Problems____.

B. Availability of Instructional Materials Such as:

Books, Workbooks, Supplies, Equipment, A.V. Aids.
Available____; Not Readily Available____.

C. Physical Facilities Such as:

Furniture, Light, Ventilation, Storage, General Conditions in Room and Building:
Adequate____, Inadequate____.

Fig. 4--"Tenure Teacher Evaluation Report," Part I, Used in Hawaii. (Continued on next page.)

TEACHER______________________________
Last, First

II. TEACHING LOAD

A. Teaching Responsibilities

1. Number of teaching periods____.
2. Total number of students enrolled____.
3. Number of subject preparations____.
4. Special Teaching Programs:
   Yes_____ No_____

B. Non-Teaching Responsibilities

1. Grade level and Committee Meetings
2. Special Assignments such as: School Programs; A.V.; J.P.O.; Yard Duty; Lunch Duty; Book Evaluation; Curriculum Work.
3. Other non-teaching duties such as: Clerical work; First Aid and Other Health Problems; Parent Conferences.

III. TEACHER PLACEMENT

Appropriate placement in accordance with his training and experience.
Yes____ No_____

IV. OTHER CONDITIONS AFFECTING TEACHER PERFORMANCE

Such as: Classroom Interruptions; Community Service; Money Raising Activities; Special Requests of Teachers

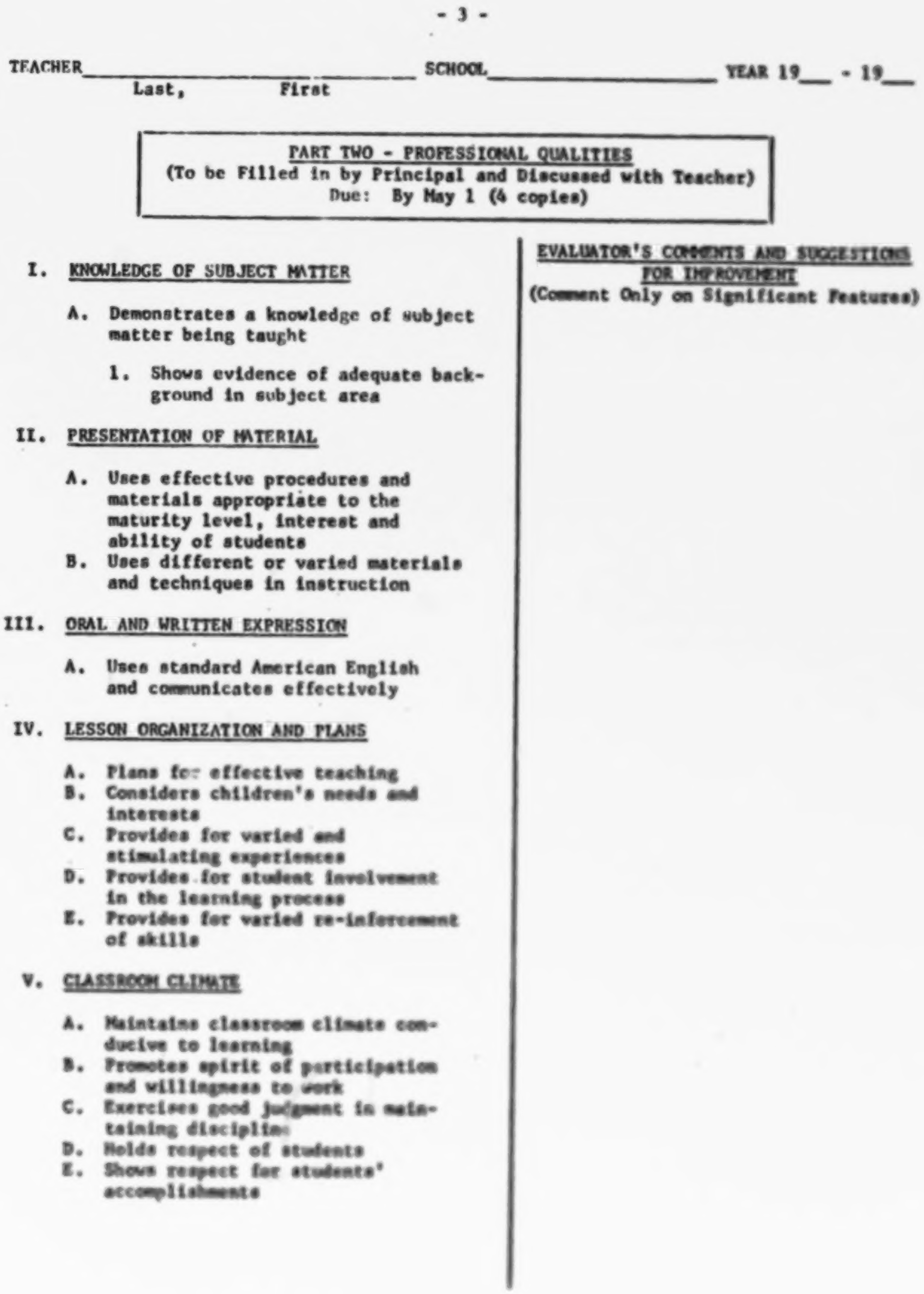
TEACHER'S COMMENTS AND SUGGESTIONS
(Comment on Items that Apply to You)

______________________________ ______________________
Teacher's Signature Date

COMMENTS BY PRINCIPAL:

______________________________ ______________________
Principal's Signature Date

RECEIVED
CERTIFICATION
AUG 18 1966
STATE DEPT. OF EDUCATION
FLORIDA



- 3 -

TEACHER______________________________ SCHOOL____________________ YEAR 19___ - 19___
Last, First

PART TWO - PROFESSIONAL QUALITIES
(To be Filled in by Principal and Discussed with Teacher)
Due: By May 1 (4 copies)

I. KNOWLEDGE OF SUBJECT MATTER

- A. Demonstrates a knowledge of subject matter being taught
  - 1. Shows evidence of adequate background in subject area

II. PRESENTATION OF MATERIAL

- A. Uses effective procedures and materials appropriate to the maturity level, interest and ability of students
- B. Uses different or varied materials and techniques in instruction

III. ORAL AND WRITTEN EXPRESSION

- A. Uses standard American English and communicates effectively

IV. LESSON ORGANIZATION AND PLANS

- A. Plans for effective teaching
- B. Considers children's needs and interests
- C. Provides for varied and stimulating experiences
- D. Provides for student involvement in the learning process
- E. Provides for varied re-inforcement of skills

V. CLASSROOM CLIMATE

- A. Maintains classroom climate conducive to learning
- B. Promotes spirit of participation and willingness to work
- C. Exercises good judgment in maintaining discipline
- D. Holds respect of students
- E. Shows respect for students' accomplishments

EVALUATOR'S COMMENTS AND SUGGESTIONS FOR IMPROVEMENT
(Comment Only on Significant Features)

Fig. 5--"Tenure Teacher Evaluation Report," Part II, Used in Hawaii. (Continued on next page.)

TEACHER____________________________________
Last, First

EVALUATOR'S COMMENTS AND SUGGESTIONS FOR IMPROVEMENT
(Comment Only on Significant Features)

VI. CLASSROOM MANAGEMENT

A. Shows good organization and preparation
B. Administers routines effectively
C. Provides for satisfactory physical classroom conditions

VII. ATTITUDES AND WORKING RELATIONSHIPS

A. Accepts constructive suggestions
B. Effectively implements the suggestions for improvement
C. Likes people and works well with students, fellow teachers, administrators and parents

VIII. ORGANIZATION OF SCHOOL DETAILS

A. Keeps adequate school records
B. Meets expected deadlines
C. Knows and follows school procedures and rules

IX. PROFESSIONAL IMPROVEMENT

A. Keeps up with professional trends and techniques in education
B. Improves himself professionally--e.g. workshops, institutes, college courses, reading, travel, research

PRINCIPAL'S SUMMARY:

______ Satisfactory   ______ Not Satisfactory

COMMENTS OF PRINCIPAL: (If additional space is needed, attach another sheet.)

COMMENTS BY TEACHER: (If additional space is needed, attach another sheet.)

(Teacher's signature does not necessarily indicate approval but merely that he is aware of evaluation)

____
Teacher's Signature

____
Date

____
Principal's Signature

____
Evaluator's Signature & Title
(If other than Principal)

____
Date

RECEIVED
AUG 1 8 19[illegible]
STATE DEPT. OF EDUCATION
FLORIDA

The staff of the Office of Personnel Services of the Hawaii State Department of Education feels that the evaluation program for probationary teachers is effective in eliminating incompetent teachers. About one percent of the new teachers are rated unsatisfactory, and consequentl are released immediately or at least are not rehired for a second year. The staff members think that the evaluation program, along with an incentive program which provides increased salaries for those teachers who voluntarily undertake a planned program of professional improvement, is effective in improving instruction in the state. They feel also that the tenure teacher evaluation program has served to stimulate concern for evaluating and improving instruction where such concern did not previously exist.

Statewide Teacher Evaluation in Florida

The specific purposes which the Florida teacher evaluation program is expected to fulfill have not been defined. The Florida statutes[1] do require, however, that each certified school employee (teacher, administrator, etc.) be evaluated annually and that this evaluation be placed on file in the State Department of Education. Two additional copies of the evaluation form are completed: one for the county office and one for the files of the local school. Evaluations of teachers are normally completed by principals and countersigned by county superintendents. In the State Department of Education, the evaluations become part of the teacher personnel files. They are available for inspection only by State Department of Education personnel and by county school personnel on official business.

---

[1] Statutes of Florida (1965), Section 231.25.

A standard form to be used for the evaluations has been adopted by the State Board of Education (Figure 6). The form is designed so that it can be read with an optical scanner. The ratings are thus transferred automatically to punched cards. The punched card data are transferred to magnetic tape for retrieval and analysis *via* computer. This has made possible the summarizing of ratings for individual schools, for counties, and for the entire state. Statewide Summaries of ratings for the years 1963-64 and 1964-65 are presented in Tables 1 and 2.

Data from the Florida teacher evaluation forms have been used by a number of Florida State University students in their graduate thesis and dissertation projects. Studies of this type which have been completed to date made use of an earlier version of the evaluation form. This version differs from the present in that it contains more items and employs a scale with five response positions, rather than three. The nature of the items on the new and old forms is the same, however.

Wurzbach[1] studied the relationship between principals' ratings and three status factors: degrees held, teaching experience and county of employment. He found a significant amount of variance in

---

[1] Edward G. Wurzbach, "A Three-County Study of Teachers' State Evaluation Scores in Relation to Their Experience and Type of Degree" (unpublished M.S. thesis, Florida State University, Tallahassee, 1962).

Fig. 6--Florida Teacher Evaluation Form

TABLE 1.--Ratings assigned to 45,035 Florida teachers for the 1963-64 school year using the Florida Teacher Evaluation Form

Evaluation Item	Number and Percent of Responses									
	Superior		Satisfactory		Poor		Yes		No	
	No.	%	No.	%	No	%	No.	%	No.	%
I. Personal Qualifications										
A. General Health and Emotional Stability	15,704	35	27,897	62	1,277	3				
B. Personal Appearance--Neat & Well-Groomed	18,862	42	25,160	56	858	2				
C. Ability to Think Logically & to Make Practical Decisions	16,432	36	26,570	59	1,852	4				
D. Punctuality	22,649	50	20,580	46	1,648	4				
E. Accuracy	19,762	44	24,169	54	915	2				
F. Ability to Take Necessary & Appropriate Action on His Own	17,822	40	25,329	56	1,694	4				
G. Professional Dedication	22,740	50	21,323	47	796	2				
II. Relationships with Others										
A. Is Respected by Pupils	22,921	51	20,770	46	1,105	2				
B. Is Responsible & Dependable	20,418	45	22,855	51	1,577	4				
C. Is Friendly, Understanding, Sympathetic with Community, Other Staff Members & Administration	22,533	50	20,918	46	1,314	3				
D. Is Morally Upright							44,505	99	301	1
E. Is Professionally Ethical							44,505	99	301	1
III. Teaching Skills & Ability										
A. Knows Subject Matter	19,911	44	24,261	54	339	1				
B. Takes Action to Improve Himself	17,251	38	26,295	58	1,165	3				
C. Uses Instructional Materials & Lesson Plans Effectively	17,922	40	25,408	56	1,270	3				
D. Develops Pupil Interest & Eagerness to Learn	18,927	42	23,814	53	1,839	4				
E. Maintains Pupil Control	18,927	42	23,814	53	1,839	4				
F. Uses Material in Cumulative Folder	10,904	24	31,006	69	1,749	4				
IV. Would You Recommend This Teacher for Re-Employment							42,887	95	1,358	4

TABLE 2.--Ratings assigned to 49,977 Florida teachers for the 1964-65 school year using the Florida Teacher Evaluation Form

Evaluation Item	Number and Percent of Responses									
	Superior		Satisfactory		Poor		Yes		No	
	No.	%	No.	%	No.	%	No.	%	No.	%
I. Personal Qualifications										
A. General Health & Emotional Stability	16,630	33	31,414	63	1,532	3				
B. Personal Appearance--Neat & Well Groomed	22,142	44	26,934	54	640	1				
C. Ability to Think Logically & to Make Practical Decisions	20,095	40	28,193	56	1,361	3				
D. Punctuality	25,944	52	22,361	45	1,395	3				
E. Accuracy	22,873	46	26,112	52	704	1				
F. Ability to Take Necessary & Appropriate Action on His Own	21,236	42	26,974	54	1,446	3				
G. Professional Dedication	25,804	52	23,053	46	841	2				
II. Relationships with Others										
A. Is Respected by Pupils	25,616	51	23,042	46	974	2				
B. Is Responsible & Dependable	28,985	58	19,635	39	1,050	2				
C. Is Friendly, Understanding, Sympathetic with Community, Other Staff Members & Administration	26,142	52	22,493	45	991	2				
D. Is Morally Upright							49,462	99	135	1
E. Is Professionally Ethical							48,605	97	967	2
III. Teaching Skills & Ability										
A. Knows Subject Matter	22,698	45	26,122	52	311	1				
B. Takes Action to Improve Himself	19,320	39	29,180	58	981	2				
C. Uses Instructional Materials & Lesson Plans Effectively	19,384	39	28,807	58	1,040	2				
D. Develops Pupil Interest & Eagerness to Learn	19,787	40	28,229	56	1,296	3				
E. Maintains Pupil Control	22,703	45	24,680	49	1,892	4				
F. Uses Material in Cumulative Folder	12,963	26	34,735	70	1,190	2				
IV. Would You Recommend This Teacher for Re-Employment							47,667	95	1,033	2

the ratings attributable to differences in collegiate degrees held. His procedures might be questioned, however, as he apparently summed the composite scores for each section of the rating sheet in order to obtain the figures which were used in the analysis. To justify such a practice, it is necessary to make tenable the assumption that all items on the form can be assigned equal weight.

Carter[1] used the state evaluation form to distinguish between high and low merit-rated junior college teachers. He then studied the relationships between these ratings and personal, educational, and experience factors. He found few significant relationships. He, like Wurzbach, also used composite scores as his measure of teacher competence.

Gerlock[2] compared the ratings given professionally and provisionally certificated first year teachers. He did separate item-by-item analyses and found that those teachers meeting professional certification standards scored significantly higher on some items, particularly those relating to teaching skill.

Four additional studies dealing with factors which might warrant consideration in interpreting data collected with the Florida evaluation form have also been completed. McTeer[3] investigated the hypothesis that teacher-principal likenesses are a factor in teacher rating. He assumed that the more similarities which existed between the teacher and the principal, the higher would be the teacher's rating. Factors which he studied include age, grade point average, highest degree held, and amount of professional work as an undergraduate. He found some significant correlations, but in no case did the likeness factors account for a substantial amount of the variance. Some shortcomings of his study are that he apparently combined scores on scale items and did not account for difference in the general level of scores assigned by different principals.

---

[1] Fletcher Fairwick Carter, "Selected Aspects of Pre-Service Preparation and Prior Experience of High and Low Merit-Rated Junior College Teachers in Florida" (unpublished Ph.D. dissertation, Florida State University, Tallahassee, 1964).

[2] Donald E. Gerlock, "An Analysis of Administrators' Evaluations of Selected Professionally and Provisionally Certificated Secondary School Teachers" (unpublished Ed.D. dissertation, Florida State University, Tallahassee, 1964).

[3] John Hugh McTeer, "A Study of the Relationship of Teacher-Principal Likenesses and the Principals' Ratings of Teacher Effectiveness" (unpublished Ed.D. dissertation, Florida State University, Tallahassee, 1963).

Packer[1] was interested in the sensitivity of the instrument to "self-others acceptance" which he deemed an essential trait for an effective teacher. He found very little relationship between this construct and the principals' ratings and concluded that either the principals were not sufficiently familiar with the teaching situations or they were not appreciative of "self-others acceptance" as a determinant of effect teaching.

Two recent studies by Daniel employed the current version of the Florida teacher evaluation form. The first employed the analysis of variance technique to determine whether certain factors could account statistically for differences in ratings assigned to a teacher on any items on the form.[2] The factors tested were (a) subject or grade being taught (i.e., Is there any difference between ratings assigned to teachers of one grade and those assigned to teachers of another grade?), (b) evaluator (i.e., Is there any difference between the level of ratings assigned by one evaluator and those assigned by others?), (c) certificate rank (i.e., education level), (d) number of different preparations which the teacher must make, (e) sex of the principal, and (f) age of the principal. The only one of these factors which contributed significantly to the variance in ratings was the evaluator effect. This can be interpreted two ways: either some principals tend to rate their teachers higher than do other principals, or some principals have better teachers in their schools than do others. A subsequent series of studies has been begun to investigate this problem further. The value of such efforts is limited, however, because technical deficiencies of the evaluation form place limitations on interpretations which can be made from statistical manipulations of the data therefrom. This problem is discussed later in greater detail.

---

[1] Morton Alfred Packer, "A Study of the Relationship Between Teachers' Self-Others Acceptance and the Principals' Ratings of These Teachers" (unpublished Ed.D. dissertation, Florida State University, Tallahassee, 1964).

[2] K. Fred Daniel, "A Catalog of Analysis of Variance Pilot Studies" (unpublished research report, Multi-State Teacher Education Project, State Department of Education, Tallahassee, Florida, 1966).

A second study employing the present form deals with the objectivity of ratings assigned by principals.[1] Objectivity is defined as the extent to which independent ratings of one teacher completed by two different evaluators are in agreement. In this study, ratings by assistant principals and by supervisors were compared with those submitted by principals. These ratings are summarized in Table 3. It was found in both the sample of assistant principals and the sample of supervisors that the principal and the other rater agree about two-thirds of the

---

[1] K. Fred Daniel, "A Study of the Objectivity of the Florida Teacher Evaluation Form" (unpublished research report, Multi-State Teacher Education Project, State Department of Education, Tallahassee, Florida, 1967).

TABLE 3.--Comparisons of ratings assigned to the same teachers using the official Florida Teacher Evaluation Form

	Comparison of supervisors' and principals' ratings						Comparison of assistant principals' and principals' ratings					
	Supervisors' ratings higher than principals'		Supervisors' ratings lower than principals'		Supervisors' and principals' ratings same		Assistant principals' ratings higher than principals'		Assistant principals' ratings lower than principals'		Assistant principals' and principals' ratings same	
I. PERSONAL QUALIFICATIONS	N	%	N	%	N	%	N	%	N	%	N	%
A. Is Healthy and Emotionally Stable	49	28	30	17	95	55	62	18	64	18	222	64
B. Is Neat and Well Groomed In Appearance	41	24	26	15	107	62	54	15	70	20	222	65
C. Thinks Logically and Makes Practical Decisions	28	16	27	16	118	68	55	16	80	23	213	61
D. Is Accurate	27	16	32	18	115	66	45	13	76	22	227	65
E. Is Punctual	38	22	40	23	95	55	56	16	74	21	217	63
F. Takes Necessary and Appropriate Action On His Own	28	16	35	20	109	63	43	12	82	24	221	64
G. Is Dedicated to His Profession	27	16	38	22	107	62	38	11	80	23	229	66
II. RELATIONSHIPS WITH OTHERS												
A. Is Respected by Pupils	43	25	29	16	102	59	44	11	70	20	232	68
B. Is Responsible and Dependable	36	21	27	16	109	63	39	11	83	24	229	65
C. Is Friendly, Understanding, Sympathetic	28	16	27	16	118	68	38	11	94	28	215	62

TABLE 3.--Continued

	Comparison of supervisors' and principals' ratings						Comparison of assistant principals' and principals' ratings					
	Supervisors' ratings higher than principals		Supervisors' ratings lower than principals'		Supervisors' and principals' ratings same		Assistant principals' ratings higher than principals'		Assistant principals' ratings lower than principals'		Assistant principals' and principals' ratings same	
	N	%	N	%	N	%	N	%	N	%	N	%
D. Is Morally Upright*	0	0	0	0	171	100	0	0	3	1	344	99
E. Is Professionally Ethical*	1	1	9	5	161	94	6	2	8	2	330	96
III. TEACHING ABILITY												
A. Knows Subject Matter	23	13	31	19	109	64	60	16	55	16	242	68
B. Takes Action To Improve Himself	26	15	42	27	99	58	53	15	66	19	231	66
C. Uses Instructional Materials and Lesson Plans Effectively	31	18	22	13	118	69	52	15	61	18	227	66
D. Develops Pupil Interest and Eagerness to Learn	35	20	26	15	111	65	57	17	73	21	216	63
E. Maintains Pupil Control	37	22	27	16	106	62	60	17	75	21	216	61
F. Uses Material In Cumulative Folder	29	18	20	13	109	69	46	14	66	19	225	68
IV. WOULD YOU RECOMMEND THIS TEACHER FOR RE-EMPLOYMENT?*	1	1	2	1	155	98	2	1	16	5	315	94
TOTAL	530	16	495	15	2213	68	826	13	1129	18	4397	69

*An item with only two possible responses: yes and no.

time. (With the distributions of ratings observed, agreement could usually be expected about half of the time by chance.)[1]

Evaluating the Florida teacher evaluation program is difficult since there is no specific purpose which the program has been designated to fulfill. It might be assumed, however, since the evaluation forms are devised for machine processing and since routines have been established to tabulate and summarize the data, that there has been some intention to secure comprehensive information relating to the quality of teaching in the state. This is information which might be used in developing ways to improve the educational program. Further evidence of such an intention can be found in the proposal for a "multi-state project to improve teacher education,"[2] submitted to the U. S. Office of Education in 1965. It was proposed that the Florida State Department of Education use its teacher evaluation data along with other data to determine needed services in the areas of pre-service and inservice teacher education.

When viewed in this framework, the Florida teacher evaluation program displays striking inadequacies. The technical deficiencies of the instrument are one example. The extreme skewing in the distribution of ratings assigned on the form limits the utilization of parametric statistical techniques when analyzing the data. Thus, inferences drawn from the data using standard scientific procedures must be interpreted with caution. The lack of homoscedacticity which can result from the skewing often makes it inadvisable to calculate Pearson product-moment coefficients of correlation to provide indices of the relationships between ratings received and other variables. The limited number of steps on the rating scale virtually eliminates the possibility of using this index.

---

[1] Probabilities of agreement between principals' ratings and assistant principals' ratings on a given item were calculated by multiplying the proportion of the sample rated in each category by principals times the proportion rated in the same category by assistant principals. These products were then summed to obtain the probability of agreement on that item.

[2] This project has been funded under Title V of Public Law 89-10 (Elementary and Secondary Education Act of 1965) and is now called the Multi-State Teacher Education Project. The proposal was submitted by the Maryland State Department of Education, the project administrator. Other states in the Project are Florida, Michigan, South Carolina, Utah, Washington, and West Virginia.

The principal handicap which is imposed by a rating scale with a limited number of steps results from restrictions upon the amount of information which the scale can provide. The optimum number of scale points is that which makes maximum use of the observer's discriminative powers. This point is reached when the ratio between true variance and error variance is maximum. Guilford[1] reviews a number of studies which deal with the optimum number of steps to be used in a rating scale. He concludes that the number is usually greater than seven and may, in certain situations, be as high as twenty-five. The Florida form employs scales with three, and in some cases two, steps.

The Florida evaluation form also appears to have some conceptual deficiencies. Particularly notable is the great emphasis placed upon general attitudes or personality traits which are deemed to be conducive to effective teaching and the small amount of emphasis upon behaviors which definitely fall into the latter category are, "Uses instructional materials and lesson plans effectively," "Develops pupil interest and eagerness to learn," and "Maintains pupil control." Following the distinction which was made in Chapter I, only these items fall within the realm of evaluation; the others constitute predictors.

---

[1] J. P. Guilford, Psychometric Methods (New York: McGraw-Hill Book Company, Inc., 1954), pp. 289-291.

To summarize, it appears that, because of deficiencies of the evaluation instrument, the Florida teacher evaluation program can provide very little information which can be used to improve education in Florida

Concluding Statement

It appears that neither a review of the educational literature nor a study of current practices can provide direct guidance in developing a program for collecting and analyzing information on teaching to facilitate state educational leadership. Only two references were discovered which make mention of the subject. Two of the three state programs for evaluating teaching which are now operating (Pennsylvania and Hawaii) are not designed for obtaining the type of data which could be used for implementing state educational leadership. The third program (Florida), which appears to have been conceived for such a purpose, displays some serious inadequacies.

Thus, to proceed in developing a system for collecting and analyzing the desired data, it is necessary to turn elsewhere for assistance. In the present study, general approaches to the evaluation of teaching are discussed. Then, a system for assembling and analyzing data obtained with these approaches is proposed.

CHAPTER III

APPROACHES TO THE EVALUATION OF TEACHING

Evaluation was described in Chapter I as the process of comparing evidence with criteria. The amount of evidence which could be gathered through the observation of teaching processes and products is overwhelming. Every syllable uttered and every movement made by teacher and pupils could be recorded on tape or film. In addition, sensing devices could be used (as has been done with astronauts in outer space) to monitor and record various physiological actions and reactions of teachers and pupils. Thus a prime goal in developing methods for evaluating teaching is to reduce the evidence to an amount which is manageable and to include in the portion retained that which is significant. Whether or not a given observation or rating system will in practice prove adequate depends upon its adaptability to the criteria for evaluation which have been adopted, for it is the criteria which determine which evidence is significant.

The criteria dictate not only the type of evidence to be sought, but also the specific approach or approaches which are appropriate for obtaining the measures which will constitute the evidence. It follows then that an approach for collecting evidence cannot be decided upon until criteria for evaluating teaching are selected.

Selecting Criteria

The types of criteria which have been used in evaluating teaching have been classified by Mitzel[1] in three categories: product criteria, process criteria, and presage criteria. Product criteria comprise the outcomes toward which teaching is directed: changes in behavior of pupils. Such changes are often declared to be the ultimate criteria of teacher effectiveness. Writers expressing this opinion include the Committee on the Criteria of Teacher Effectiveness appointed by the American Educational Research Association,[2] Ackerman,[3] and Yamamoto.[4]

It is maintained by many persons, however, that factors other than the influence of the teacher contribute significantly to changes in pupil behavior and, thus, it is not possible to evaluate the work of a teacher solely in terms of the achievement of his pupils. In an effort to account for extraneous influences and still use changes in pupils as a criterion, modified approaches for deriving pupil gain scores have sometimes been used. These include achievement quotients, residual gain scores, and scores obtained through variations on the residual gain procedure.[1] However, in a large portion of the research dealing with teaching effectiveness, product criteria have not been employed. Medley and Mitzel state that the proportion of studies in which measures other than pupil growth have been used as criteria approximates 90 per cent.[2] Only nineteen out of 138 studies in a summary compiled by Barr[3] employed pupil gain as a criterion.

---

[1] Harold E. Mitzel, "Teacher Effectiveness," *Encyclopedia of Educational Research*, ed. Chester W. Harris (New York: The Macmillan Company, 1960), pp. 1482-85.

[2] H. H. Remmers, *et al.*, "Second Report of the Committee on Criteria of Teacher Effectiveness," *Journal of Educational Research*, Vol. 46 (May, 1953), p. 642.

[3] Walter I. Ackerman, "Teacher Competence and Pupil Change," *Harvard Educational Review*, Vol. 24 (Fall, 1954), p. 274.

[4] Kaoru Yamamoto, "Evaluating Teacher Effectiveness: A Review of Research," *Journal of School Psychology*, Vol. 2 (January, 1963), p. 61.

Process criteria are those types of teacher behavior believed to be desirable, at least in given situations. They constitute an obvious alternative for consideration when problems of securing and interpreting information on pupil gains are deemed sufficiently serious to preclude the use of product criteria. In the domain of the local administrator or other instructional leader (as contrasted with the domain of the researcher),[4] process criteria are of particular significance. Here is where decisions relative to retention, promotion, salary, supervision, inservice education programs, and other types of staff development must be made. These are ongoing decisions which cannot wait until evidence becomes available which would permit teaching to be evaluated in terms of ultimate achievements of pupils.

---

[1] William J. Ellena, Margaret Stevenson, and Harold Webb (eds.), Who's a Good Teacher? (Washington, D. C.: American Association of School Administrators, 1961), p. 15.

[2] Donald M. Medley and Harold E. Mitzel, "The Scientific Study of Teacher Behavior," Theory and Research in Teaching, ed. Arno A. Bellack (New York: Bureau of Publications, Teachers College, Columbia University, 1963), p. 83.

[3] Arvil S. Barr, "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, Vol. 16 (June, 1948), pp. 203-283.

[4] Ellena, Stephenson, and Webb (eds.), pp. 5-6.

Presage criteria are made up of those traits and experiences which are thought to be fundamental to certain facets of teaching performance (which can be evaluated employing process criteria) or to achieving certain outcomes (which can be evaluated employing product criteria). Thus, presage criteria function primarily as predictors. Their value is determined by their efficiency in predicting which persons will teach effectively and which will not. Studies have, nevertheless, been undertaken employing presage variables as criteria, rather than predictors. Also, it is not uncommon to find on rating forms items such as, "Is healthy," or "Is well versed in subject matter." These items relate to traits which are commonly thought to be antecendents of creditable teaching. They do not, in themselves, comprise measures of teaching.

The various specific factors which can be incorporated into criteria for evaluating teaching are numerous. The product, process, and presage categories each encompass a multitude of possibilities. The Taxonomy of Educational Objectives[1] illustrates several categories which might be used in developing product criteria. Since measures in each of the categories can be taken after varying lapses of time, the potential number of specific product criteria which can be adopted is compounded immediately.

---

[1] Benjamin S. Bloom (ed.), Taxonomy of Educational Objectives, Handbook I: Cognitive Domain (New York: David McKay Company, Inc., 1956); David R. Krathwohl, et al., Taxonomy of Educational Objectives, Handbook II: Affective Domain (New York: David McKay Company, Inc., 1964).

Process criteria can also encompass several different dimensions. Brown[1] has developed instruments to measure the extent to which practices of teachers agree or disagree with practices endorsed by the John Dewey philosophy. Flanders[2] has devised a series of categories to measure the nature of the verbal influence which teachers attempt to impose on pupils. His categories consist of types of direct or indirect influence. "Using student ideas" is an example of indirect influence and "giving directions" is an example of direct influence. Smith[3] has devised a system for analyzing in different terms the teacher's use of language. His categories comprise "logical operations;" defining, describing, and conditional inferring are examples. A group at Stanford University[4] has analyzed teaching into several separate sub-skills or tasks and has devised criteria for evaluating each of these. Examples of the skills are making assignments, monitoring in-class assignments, small group work, and introducing a unit.

---

[1] Bob Burton Brown, "Bringing Philosophy into the Study of Teacher Effectiveness," Journal of Teacher Education, Vol. 17 (Spring, 1966), pp. 35-40.

[2] Edmund Amidon and Ned A. Flanders, The Role of the Teachers in the Classroom: A Manual for Understanding and Improving Teachers' Classroom Behavior (Minneapolis: Paul S. Amidon and Associates, 1963).

[3] B. Othanel Smith and Milton Meux, "A Study of the Logic of Teaching," Cooperative Research Project Number 258, Trial Edition (Urbana, Illinois: Bureau of Educational Research, College of Education, University of Illinois, 1962).

[4] James M. Cooper, "A Performance Curriculum for Teacher Education," The Second Annual Florence B. Stratemeyer Lecture read before the Annual Meeting of the Association for Student Teaching, Chicago, February 16, 1967.

Presage criteria can also include a multitude of different factors. The range is restricted only by such limits as might be inherent in the ability of man to categorize human traits and experiences.

Are Some Criteria More Worthwhile than Others?

With so many possible criteria available, it is reasonable to ask whether some criteria might not possess intrinsically greater merit than others. Are there guidelines which might help professional educators to determine which criteria would be most acceptable in a given situation?

According to Rabinowitz and Travers, there are available no objective procedures for identifying acceptable or unacceptable criteria of teaching effectiveness. Effective teaching does not exist independently but is an artifact created when an independent or collective value judgment is made. The authors state that, "No teacher is more effective than another except as someone so decides and designates. . . . The ultimate definition of the effective teacher does not involve discovery but decree."[1] Rabinowitz and Travers suggest that the place to start in developing criteria is with the goals which the teaching is expected to accomplish. The teaching which contributes to the attainment of these goals is thus considered effective. The goals are, of course established on the basis of value judgments. Travers and Rabinowitz eventually go so far as to declare that pupil growth is the most sensible criterion of effective teaching. However, this recommendation must be considered in light of their earlier statement.

---

[1] William Rabinowitz and Robert M. W. Travers, "Problems of Defining and Assessing Teacher Effectiveness," Educational Theory, Vol. 3 (July, 1953), p. 212.

Ryans[1] also believes that no type of criterion of effective teaching possesses intrinsic goodness. The worthiness of any given set of criteria is dictated by the values of the specific culture which the teaching is intended to serve.

Yildirim,[2] however, maintains that it is an overstatement to say that there is no basis for validating a criterion outside of what someone decrees to be important. He feels that such arbitrariness serves to affirm that the judgment of one person is as good as that of any other person: that the judgment of the layman is equally as valid as that of the professional.

Yildirim's position is that knowledge and understanding of the situation in which the criteria are to apply is necessary in order to arrive at the optimum criteria. He admits that the criteria result from a value judgment, but this judgment must also have an empirical base. The persons establishing the criteria must maintain contact with the realities of the teaching situation. Criteria without such a base are likely to be capricious.

---

[1] David G. Ryans, Characteristics of Teachers: Their Description, Comparison and Appraisal (Washington, D. C.: American Council on Education, 1960), p. 16.

[2] Cemal Yildirim, "An Analytic Model for Evaluation of Teacher Competence" (unpublished Ph.D. dissertation, Indiana University, Bloomington, 1963), pp. 95-102.

It was reasoning along the lines of the Yildirim argument that led to the adoption of one of the guidelines set forth in Chapter I of the present study: the one declaring that the state should not dictate to professional educators the procedures which they should follow or the specific outcomes which they should seek in their efforts to implement local educational objectives. Recommendations regarding evaluative procedures or criteria should be mediated by knowledge resulting from familiarity with the situations to be affected by the recommendations.

Some Clarifications Regarding the Present Study

It was specified in Chapter I that this study deals with measurement and evaluation of teaching but does not deal with the problems of predicting the quality of teaching which might be expected in any given circumstance. Therefore, it does not deal with presage criteria since their relevance is principally in the area of prediction. Because of this, systematic approaches for measuring antecedents of teaching (i.e., presage or status factors) are not treated in the present chapter. It should be noted, however, that if the traits which comprise presage criteria are described in behavioral terms and thus can be observed in the classroom, the presage criteria at once become process criteria. Hence, the distinction between presage and process criteria is not as decisive as it might appear at first encounter.

Likewise, the distinction between process and product criteria is not always obvious unless everything that the teacher does is considered process and everything which the pupils do is considered product. However, when pupils at the direction of the teacher line up on two sides of the room for a spelling contest, it seems more reasonable to consider this activity as part of the teaching process and to consider the degree of accuracy with which the pupils spell the words as part of the teaching product.

The distinction is not critical, however. The point is that criteria can relate to things that happen before the teacher begins to teach, to things that happen in the course of a teacher's teaching, and to things that happen after the teaching is completed. This chapter deals with procedures for evaluating teaching in terms of those things which happen during or after a teacher's teaching which, on the basis of the adopted criteria, are considered significant.

Collecting Evidence and Comparing it with Criteria

As the above discussion indicates, the criteria for evaluating teaching which can be selected comprise a varying assortment. Likewise, several different approaches which can be employed to gather evidence and to compare it with criteria are available. Not all of these approaches are equally applicable to any given type of criterion. Also, not all are equally suitable for assembling information which can be used in implementing state education leadership. The remaining portion of this chapter contains descriptions of a diversity of approaches. The descriptions are intended to provide an indication of the range of choices available. The utility of various approaches, in terms of their potential for contributing information for implementing state educational leadership, is treated more explicitly in Chapter IV.

In the subsequent discussion, procedures for collecting evidence and comparing it with criteria are conceived as varying along three dimensions. The dimensions represent (1) the situation in which the evidence is collected, which varies from a "normal" situation to a "constructed" situation, (2) the agent responsible for coding and recording the information, which has "self" (i.e., the teacher or pupil performing) at one extreme and "others" at the opposite extreme, and (3) the relationship between collecting evidence and comparing evidence with criteria (i.e., evaluating), which can vary from "evaluation while observing" to "evaluation independent of observation." These dimensions are described below in greater detail. They are also depicted graphically in Figure 7. Any evaluative procedure can be classified in the three-dimensional space of this diagram. The letters A, B, C, D, E, F, G, and H identify extreme points on each of the dimensions in the three-dimensional space. Following further explanation of the dimensions, approaches to evaluation are discussed and couched in terms of these extremes. It should be kept in mind, however, that there are also infinite intermediate positions between the extremes.

The first dimension relates to the situation in which the evidence is collected. At one pole of this dimension is the normal day-to-day teaching situation or the normal life situation in which an individual might be found subsequent to being taught. The latter type of situation would be of interest if teaching were being evaluated in terms of certain types of product criteria. At the opposite pole of this dimension is a situation constructed to elicit certain types

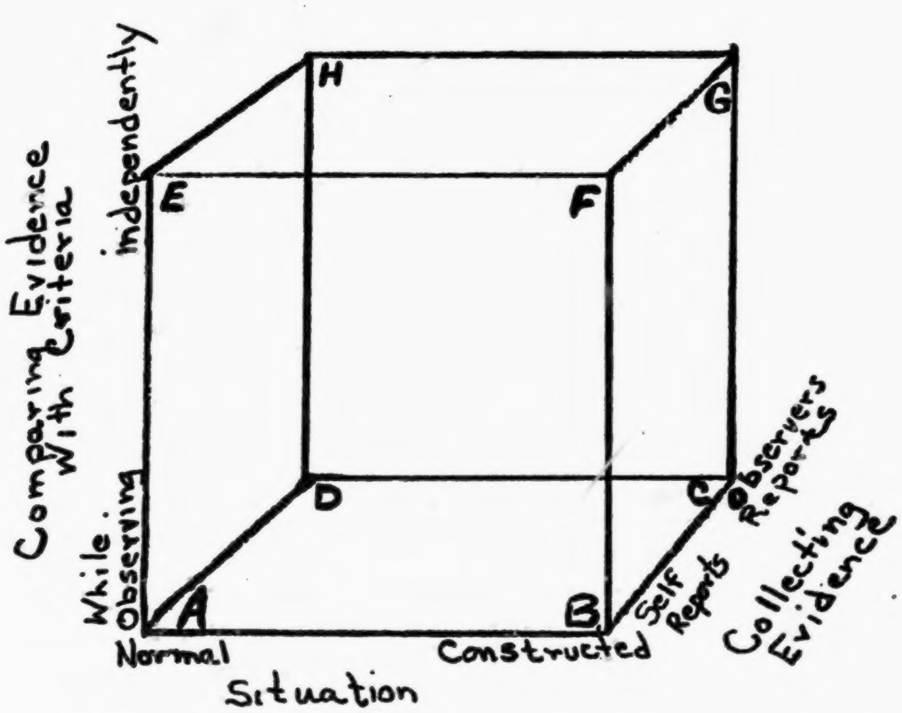

Comparing Evidence With Criteria
Independently
While Observing
H
G
E
F
D
C
A
B
Normal
Constructed
Situation
Self Reports
Observers Reports
Collecting Evidence

Fig. 7--Graphic representation of system for classifying evaluation procedures along three dimensions.

of specific evidence. A test or examination is an example of a constructed situation. This, of course, represents the approach normally employed to measure pupil learning, the desired product of teaching.

Constructed or standardized situations are seldom, if ever, used in evaluating the teaching performance of inservice teachers. It seems possible, however, that if criteria for effective teaching were stated in terms of specific teaching tasks (as described by Cooper[1] for example) that standardized situations could be constructed to measure objectively and to evaluate the performance of teachers on these tasks. The classroom simulator designed by Kersh[2] might serve as a prototype for a mechanism which could be used in presenting stimuli in the constructed situation.

---

[1] Cooper.

[2] Bert Y. Kersh, "Classroom Simulation: A New Dimension in Teacher Education," Title VII Project Number 886, National Defense Education Act of 1958 (Monmouth, Oregon: Teaching Research, Oregon System of Higher Education, 1963).

The second dimension relates to the agent who codes the evidence. At one pole, the evidence is coded by the person performing the behavior. This would be the teacher if the evidence relates to teaching performance, or the pupil if the evidence relates to teaching products. In the usual objective test, the established procedure for responding is designed so that the responses are recorded in coded form by the person taking the test. In an essay test or a performance test of almost any type (e.g., a music contest, an ice skating contest, a debate tournament, or a job interview), the evidence which will be compared with the evaluation criteria is coded by someone other than the performer.

The third dimension relates to the point at which the actual evaluation takes place; that is, the point where the evidence is compared with the criteria. At one pole of this dimension, the evaluation takes place during the observation process. At the opposite pole, the collection of the evidence and the evaluation of the evidence are carried out independently. With rating procedures, the recorded rating constitutes an evaluation. In other words, the person making the observation compares what he sees with the applicable criteria and then records the results of the comparison.

When the collection of evidence and the evaluation of the evidence are carried out independently, a procedure is developed whereby the observer describes, in narrative or quantitative terms, what has taken place. This description is limited to those aspects of the performance of the teacher and/or pupils which are relevant to the criteria. An example of this approach is the interaction analysis system which has been adopted to the classroom by Flanders[1] and others. The evaluation of the teaching takes place after the observation record has been compiled. With such a system, the evaluation can be done by persons other than those who conducted the observation. The same data can also be analyzed employing different criteria, providing, of course, that the alternate criteria encompass the same elements which were recorded in the observation.

Self Reports and Self Evaluation

Self evaluation is the fundamental procedure by which teachers can improve their teaching. Considering self reporting and self evaluation in terms of process criteria, area A in Figure 7 represents the self evaluation carried out by individual teachers in day-to-day teaching situations. Area E represents a procedure whereby self reports are prepared by a teacher and evaluated independently, possibly with the aid of a consultant. This type of practice seems most likely in a situation where a teacher is having a special problem and seeks outside help. Areas B and F parallel A and E except that the measures are taken in a constructed or standardized situation. Such a situation would be possible only in an institution which has provisions for simulating classroom situations.

If teaching were being evaluated in terms of product criteria, self reports could be solicited from pupils. This would take place in a constructed situation unless it happened that the desired data were emitted spontaneously or as part of some other activity. The data obtained from the pupils could be in the form of judgments (areas A and B) or in the form of descriptive information or responses to test questions (areas E and F). The latter type of information would be evaluated independently.

---

[1] Amidon and Flanders.

In general, self reports from teachers which describe or evaluate the teaching which has taken place in their classrooms would be of limited use in implementing state educational leadership. According to Biddle, the cognitive systems normally used by teachers for thinking about classroom activities are not adequate for objectively describing classroom situations.[1] It seems, however, that, if the teacher were sufficiently familiar with the criteria for evaluation, this need not be the case. Nevertheless, the problem of assuring the reliability or objectivity of the evidence remains. As Travers points out, an individual observing himself is a biased observer.[2] The self-report approach need not be eliminated categorically, though, as there might arise certain situations in which such an approach might prove efficacious, possibly when used in conjunction with other procedures.

---

[1] Bruce J. Biddle, "The Integration of Teacher Effectiveness Research," Contemporary Research on Teacher Effectiveness, ed. Bruce J. Biddle and William J. Ellena (New York: Holt, Rinehart and Winston, 1964), pp. 272-8.

[2] Robert M. W. Travers, An Introduction to Educational Research (2d ed; New York: The Macmillan Company, 1965), p. 272.

For evaluating teaching in terms of product criteria the testing approach (area F) is normally employed. This can be classified logically as another form of self-reporting. Since this is the approach typically used for evaluating pupil achievement, it is reasonable that data so obtained be adapted for evaluating teaching. Testing is discussed at greater length in a subsequent section of this Chapter.

Rating and Rating Scales

The most singular characteristic of rating as an evaluation method is that the evidence is compared with the criteria during the observation process and only the comparison (i.e., the rating) is recorded. Thus, rating is represented in Figure 7 at points A, B, C, and D. At points A and B, the rating is performed by the person whose performance is being measured. Thus, if process criteria were being employed, the teacher would perform a self evaluation and if product criteria were being employed, the pupil would evaluate his own skills. In areas C and D, the rating is assigned by an observer. Area D represents the usual situation when process criteria are employed, in which an observer comes to the classroom and rates the teaching performance. Area C represents the same type of rating but the teaching performance takes place in a standardized or constructed situation. If product criteria were employed, an observer would rate the performance of pupils in a special test situation (area C) or in their normal environment (area D).

Rating scales are by far the most widely used devices for evaluating teaching performance for both research and administrative or supervisory purposes. At least in the case of administrative and supervisory situations, this condition is likely to persist. The evidence which must be reviewed to determine whether or not teaching is effective is invariably extensive and subtle with numerous complexities which are difficult to catalog in advance. Because of this, quasi-mechanical or objective procedures for evaluating teaching have usually not been considered acceptable, and raters themselves have been required to perform the function of reducing the data to that which is significant and comparing this evidence with the relevant criteria. As a result, summarizing and processing of evidence takes place entirely within the "black box" and only conclusions are available for scrutiny. Hence, neither the data reduction process nor the evaluation process can be examined. If a case developed in which two "experts" evaluating the same teaching provided different evaluations, it would be a matter of speculation as to whether the discrepancy resulted from their selecting different evidence to process or from their applying different criteria in evaluating (unless, of course, the evaluators were available for questioning).

Fortunately, the evaluation process employing rating scales need not be so mercurial as the foregoing implies. Travers[1] suggests that the stability of results so obtained can be controlled, thus making the results meaningful by controlling both the type and quantity of information to be processed and the processing itself. This can be done by providing sufficient descriptive material with the rating form to orient the user, by constructing a rating instrument composed of specific rather than general scales, and by constructing the individual scales carefully. Practices recommended by Travers which should result in better scales include (1) defining with precision several points on each scale, (2) restricting each scale to a limited range of well-defined behavior, (3) varying the end of the scale which represents "good," and (4) avoiding the use of words such as "average" to represent the middle range of a scale. Discussions of technical considerations in rating scale development and the literature relating to their use are presented by Guilford[1] and Remmers.[2]

---

[1] Ibid., pp. 222-224.

An obvious technique for improving the reliability of ratings involves the training of raters. Such training could consist of a thorough orientation into the type of evidence which is to be considered significant and the type of criteria which are to be employed in analyzing it. This would be followed with practice in employing the scale including opportunities for comparing and discussing the ratings assigned. Practice sessions can be repeated until the desired level of reliability is reached. Garrison, who has been involved in two large-scale research projects employing a rating form for appraising teaching performance, obtained reliability coefficients of .73 to .85 with a training period in which trainees observed and evaluated four lessons[1] and coefficients of .53 to .89 with a training period in which trainees observed and evaluated only one lesson.[2]

---

[1] J. P. Guilford, "Rating Scales," in his Psychometric Methods (New York: McGraw-Hill Book Company, 1954), pp. 263-301.

[2] H. H. Remmers, "Rating Methods in Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally and Company, 1963), pp. 329-378.

In employing information collected with rating forms to implement state educational leadership, data obtained with any carefully developed form administered with carefully established procedures is acceptable, as long as the form utilizes normative, rather than ipsative, scales. Normative scales make comparisons between individuals; that is, they use a norm as their model. Ipsative scales make comparisons within individuals; that is, they compare one aspect of the performance of one individual with other aspects of the performance of that same individual. Data obtained with ipsative scales are not intended to be grouped. The combining of ratings from several different observations of several different teachers would, of course, be necessary in deriving information intended to affect statewide policies.

---

[1] Harry L. Garrison, "Evaluation of Teaching and Learning," (unpublished Ed.D. disseration, Stanford University, Stanford, California, 1964).

[2] F. Herbert Hite, "Effects of Reduced Loads and Intensive Inservice Training Upon the Classroom Behavior of Beginning Elementary Teachers," Cooperative Research Project No. 2973 (Olympia, Washington: Washington Superintendent of Public Instruction, 1966).

Separating the Observation Process from the Evaluation Process

In recent years research on teaching has been moving away from the study of teaching effectiveness as a global concept. Rather than rating teachers along an effectiveness-ineffectiveness scale and then attempting to find the correlates of these ratings, researchers have been studying relationships between several variables associated with teaching. These consist of traits of teachers, dimensions of teaching performance, characteristics of teaching situations, traits of pupils, and dimensions of pupil performance. The approach has been first to describe objectively and to measure these variables and then to look for relationships which exist between them, whether or not these relationships bear any immediate direct connection to any specific concept of effective teaching. The relationships studied are ones which may have been confounded when a more general approach to research on teaching was followed. The general assumption underlying this newer approach is that when interrelationships between variables affecting teaching are understood better, teaching will be understood better and the various combinations of conditions which contribute to effective teaching will be more easily recognized.

An outgrowth of this movement has been the development of observation procedures which can be used for describing objectively various aspects of teaching processes or products. These make it possible to separate the collecting of evidence on teaching from the evaluating of that teaching. This approach offers the potential for more objective evaluations by enabling the observer to concentrate more directly on collecting evidence and by making the evidence available for review both before and after the process of comparing it with the criteria has been carried out. This approach also allows more than one set of criteria to be applied to a given set of evidence and allows different elements of the evidence to be combined and studied in different ways.

In Figure 7, the approach in which observations are separated from ratings is represented in areas E, F, G, and H. Area H represents the situation in which an observer would collect data on the performance of a teacher or a pupil employing a standardized observation procedure; the data would be evaluated subsequently. Area F represents a constructed situation, most likely to occur when evaluating pupil learning, in which the pupil records his coded responses; these will be evaluated later. Area G represents a constructed situation with someone other than the performer recording the responses and area E represents a situation in which a performer records data on his own performance and evaluates it later.

The first step in developing an observation system is to determine what aspects of teacher or pupil performance are of interest. The second step is to categorize these elements so that they can be objectively reported by an observer or to locate a category system already in existence which can be adapted to a given evaluation program. Several such systems are already in use. The Classroom Interaction Newsletter[1] conducted a survey of category systems in use and received 315 questionnaires from persons who are using them in their research or teaching activities. Fifty-three different observational systems were represented among these. The most widely used systems were the Observation Schedule and Record (OScAR)[1] by Medley and Mitzel and the interaction analysis system developed by Flanders.[2]

---

[1] Classroom Interaction Newsletter, Vol. 1 (May, 1966), pp. 1-2.

Medley and Mitzel,[3] in a general article dealing with observation methods, described two different types of observation schedules. The first is called a category system. It consists of an exhaustive list of categories all dealing with one aspect of behavior. Every unit of behavior observed is classified into one of the categories. The completed observation record shows the total number of behavior units observed and the number classified in each category. The Flanders interaction analysis system is an example of the category type. It includes seven categories which describe the types of verbal influence exercised by the teacher. They are (1) accepts feeling, (2) praises or encourages, (3) accepts or uses ideas of student, (4) asks questions, (5) lectures, (6) gives directions, (7) criticizes or justifies authority. The first four constitute indirect influence and the final three constitute direct influence. There are two additional categories which deal with student talk, one for student talk in response to the teacher and one for student initiated talk, and a final category called silence or confusion to account for everything not covered previously. The Flanders system is designed so that not only the number of behavior units falling into each category can be studied but also so that the sequence of categories can be analyzed. The Flanders approach need not be restricted to the Flanders categories, but may also be used with such other categories as might be relevant in a given situation.

---

[1] Donald M. Medley and Harold E. Mitzel, "A Technique for Measuring Classroom Behavior," Journal of Educational Psychology, Vol. 49 (April, 1958), pp. 86-92.

[2] Amidon and Flanders.

[3] Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching, pp. 288-303.

The second approach to constructing an observation schedule described by Medley and Mitzel is called the sign system. With this system, a list of behaviors which may or may not occur is compiled. The observer then tallies those behavior units observed which meet the category definitions. It is not assumed that all behaviors which occur will be recorded. An example of this procedure is found in the Teacher Practices Observation Record[1] which is designed to determine the extent to which the practices of a teacher coincide with those advocated by the experimental philosophy of John Dewey. To use the schedule, a thirty-minute observation period is divided into three ten-minute segments. The observer checks which of the sixty-two listed teaching practices occur in each of the three segments.

The observation procedure developed by Medley and Mitzel for use with the OScAR combines the category and sign system. A

---

[1] Bob Burton Brown, Teacher's Classroom Behavior (Gainesville, Florida: Teacher Competence Research Project, College of Education, University of Florida, undated). (This is a group of instruments for use in evaluating a teacher.)

thirty-minute observation period is divided into six five-minute segments. The first, third, and fifth segments are spent tallying on a list (i.e., using the sign system) the types of activities, groups, materials, and behaviors which are observed. During the second, fourth, and sixth segments, the expressive behavior of the teacher is recorded using a category system. The categories are (1) non-verbal pupil-supportive, (2) pupil-supportive, (3) problem-structuring, (4) miscellaneous, (5) directive, (6) reproving, and (7) non-verbal reproving. The type of subject dealt with in each of the five-minute segments is also recorded.

The category system offers the advantage of accounting more thoroughly for behavior along a given dimension. To employ it, however, the number of categories must be limited so that the observer can keep them all in mind simultaneously and categorize observed behavior instantly. The sign system allows for a wider range of behaviors to be included. It does not, however, provide information as to the relative frequency of the behaviors. Both systems are applicable to programs for the evaluation of teaching employing either process or product criteria, provided, of course, that the relevant behaviors are defined and included in the list of categories used.

The training of observers is important with sign and category observation systems just as it is with rating systems. Therefore, the remarks about training of observers which were made earlier apply here also.

Testing

If product criteria are to be used in evaluating teaching, the place of testing is obvious. On the other hand, if process criteria are selected, the value of constructed situations for evaluating teaching is virtually unexplored. There seems a possibility, however (as stated earlier), that if criteria for effective teaching were stated in terms of specific tasks, standardized situations could be developed to sample the behavior of teachers in performing the individual tasks.

Areas B, C, F, and G in Figure 7 represent constructed or test situations. In area B, the performer (either teacher or pupil) prepares a self-report while performing. In area C, the observer evaluates the performance using a rating scale. In area F, the person being tested codes evidence for evaluation later; this is the normal objective testing situation. In area G, an observer records evidence for evaluation later. Areas C and G represent the usual performance testing situation.

Teachers and administrators are well aware of procedures for examining pupils to assess their learning. However, they have been generally unwilling to use the results of these assessments as a basis for evaluating teaching for several reasons which have been previously discussed. Their reasons might be summarized with two statements: (1) there are many factors which act before, during, and after a teacher's teaching which affect the amount of learning which takes place within any given individual, and (2) the tests which are available may not represent the full range of objectives toward which the educational program is directed.

The second criticism has been answered to a certain extent by the two volumes of the Taxonomy of Educational Objectives[1] which demonstrate that examination items can be written to yield information in areas which have previously received little attention. An approach for dealing with the first problem has been presented by Daniel.[2] In its present form, such merit as the idea might possess is of theoretical value only, because of the problems inherent in implementing the proposal. The general idea is to determine whether pupils taught by a given teacher have achieved as much as they could reasonably be expected to achieve. The procedure consists of first developing the educational goals or criteria and stating them in operational terms. These goals would be assigned weights based upon their judged importance. (The combined value of all weights would sum to one.) This would mean that the teacher would have in mind definite outcomes for each pupil which he teaches and also be aware of the relative importance of each. Then, on the basis of all available information (past performance, aptitude, current motivation, etc.) an estimate of the amount of achievement which could be expected for each pupil would be recorded. At the appropriate time subsequent to the teaching, the achievement of the pupil in all relevant areas would be measured and compared with his predicted or expected achievement. The ratio of each observed achievement to each expected achievement would be calculated and multiplied by its respective weight. The weighted rating would then be summed to provide an index of teaching effectiveness. Thus, a teacher whose pupils achieved as much as they were expected to achieve would earn an index of 1.00. A teacher whose pupils achieved more than they were expected to achieve would have an index greater than 1.00. A teacher whose pupils achieved less than they were expected to achieve would have an index of less than 1.00. The principal difficulty in applying such an approach to evaluating teaching would no doubt come in trying to arrive at valid estimates of expected achievement.

---

[1] Bloom; Krathwohl.

[2] K. Fred Daniel, "Effective Teaching and Educational Values," paper read before the Eighteenth Annual Meeting, Philosophy of Education Society, Southeastern Region, Tuscaloosa, Alabama, February 4, 1966.

Concluding Statement

This chapter has set out (1) to indicate the primacy of criteria in establishing an evaluation program, (2) to indicate the wide range of choices which can be considered in adopting criteria of effective teaching, and (3) to indicate the range of approaches available for collecting evidence and evaluating teaching. The fact that well conceived teacher evaluation programs may exist in the schools of a state does not, however, mean that the state is automatically provided with the information which it needs to evaluate its activities in teacher education, certification, and accreditation, and to begin immediately to exercise more effective educational leadership. A system must first be developed for collecting and analyzing the information which sound local evaluation programs can provide. An approach for doing this is outlined in Chapter IV.

CHAPTER IV

AN APPROACH TO ORGANIZING AND ANALYZING INFORMATION ON TEACHING TO IMPLEMENT STATE EDUCATIONAL LEADERSHIP

As stated in Chapter I, the purpose of this study is to delineate the conditions under which statewide programs for measuring and evaluating teaching can yield the most useful information for implementing state educational leadership. At the most general level, there are three such conditions which apply: the information must be relevant, interpretable, and reliable.

Providing for Relevance, Interpretability, and Reliability

Relevance refers to the extent of the relationship existing between the information collected and the circumstances which state educational leadership is intended to affect. In general, this leadership is directed toward helping local education authorities to establish educational programs which meet the needs of their citizenry and toward attempting to bring about conditions whereby the objectives of these programs are most likely to be attained. Thus, the actual relevance of information on teaching is best determined by the personnel in the local schools, except in areas in which the state specifies certain teaching processes to be followed or certain teaching products to be sought. However, states generally do not impose such specifications, even though they may be endorsed by special interest groups. The usual position of educational leaders on this matter has been the one represented in the guidelines set forth in Chapter I of the present study; namely, that the state, when exercising leadership, should neither pre-empt local initiative in determining educational objectives, nor dictate to professional educators the procedures which they should select to accomplish these objectives. This is in line with the concept of state educational leadership presented by Thurston and Roe in their book on State School Administration.[1] They feel that state departments of education can be more effective if they place their greatest emphasis upon techniques other than regulation.

To provide for the maximum relevance in terms of types of performance and types of results expected from teachers within the schools of the state, it is proposed that a statewide program for evaluating teaching use as its base data collected through evaluation programs designed to meet local needs. If these programs are carefully developed in terms of local educational goals, the resulting data are likely to be more relevant than data obtained from any other source.

In striving for relevance in educational research, the problem of rigor, which is a prerequisite for interpretability, is often encountered. Such is the case in the present situation. Attempting to combine data collected in different places under different conditions is certain to bring about problems in interpretation. These are discussed below. Also pertinent are remarks later in the chapter relating to pilot programs for developing exemplary evaluation methods and methods for disseminating information.

---

[1] Lee M. Thurston and William H. Roe, "The Growing Edge," in their State School Administration (New York: Harper and Brothers, 1957), pp. 402-419.

While the proposed approach admittedly contains inherent difficulties, the drawbacks of the alternatives seem even more critical. The alternatives would be either to require local school officials to collect information specifically for the state program, or for the state to employ special personnel to go into the schools and collect the information. With either approach, the nature of the information collected would still be determined by local educational goals and policies, if the guidelines stated in Chapter I are not to be violated. However, with these approaches, the initiative for evaluating teaching would be taken from the local personnel and assumed by the state. This is contrary to the philosophy of state educational leadership which has been endorsed in the present study. If the information were collected for the state by local school personnel, it seems likely that the evaluation might become a perfunctory administrative task. (There have been informal reports that the present Florida teacher evaluation program has been treated in some schools as just such a chore.) Because of the inherent difficulties in evaluating teaching, such an approach is apt to provide data which lack reliability. If, on the other hand, a staff of state evaluators were employed, it might appear to many that education had taken a step backward with the return of the state school inspector. The reception accorded such an official would certainly not be favorable.

Thus, to reiterate, the most reasonable approach for obtaining relevant data on which to base a program for utilizing information on teaching to implement state educational leadership is to make use of data collected through carefully developed local evaluation programs. This practice would assure that data obtained were relevant to the purposes toward which the local educational programs are directed.

Using data from local evaluation programs would not, however, preclude the possibility of conducting analyses which do not relate directly to the criteria adopted in individual schools. It would be possible to conduct pilot tests of numerous hypotheses by extracting, transforming, combining, or otherwise modifying the data collected. One important type of extra analysis would relate to the performance of graduates from different kinds of teacher education programs. The data could also be used in conjunction with data collected from other sources. Some of the types of analyses which could be conducted are described by Medley and Mitzel in their discussion of possibilities for making maximum use of data collected in natural (as opposed to experimental) situations.[1]

If a state program for measuring and evaluating teaching relies on data collected in individually developed local programs, difficulties are almost certain to develop when it comes time to combine data from different schools. This problem can be classed in the realm of interpretability, the second condition necessary in a state program to provide information on teaching. Interpretability refers to conditions which allow the data to be organized and analyzed in ways which will yield information that can be used for desired purposes, in this case, for the implementation of state educational leadership. It is thus apparent that relevance is a necessary, but not a sufficient, condition for interpretability. As previously stated, the information collected from local school districts would be relevant as it would represent the conditions which, in the judgment of the professional personnel of the individual schools, are basic to accomplishing the goals of the individual school programs. The presence of relevance would not guarantee, however, that the measuring scales and observation methods used in different schools would be designed in a manner which would permit data from different evaluation programs to be combined.

---

[1] Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally and Company, 1963), pp. 321-325.

On the other hand, the idea of imposing conditions which would allow for the combining of data from unique locally developed programs for evaluating teaching seems antithetical to the position developed under the discussion of relevance which encourages the exercise of local prerogative. Nevertheless, it seems unlikely that every program developed will be completely unique. There will certainly be common elements running through many of them. Once local school districts have developed their criteria, it would be possible for the state to exercise leadership in surveying, analyzing, and comparing these criteria. The common elements could be identified. The state could encourage communication between schools which use similar approaches and possibly could support or sponsor workshops for training observers in which representatives of several schools would participate. Thus, the state would not be discouraging local initiative, but would be helping schools to do a better job of evaluating teaching, employing the criteria which they deemed significant.

As evaluation programs develop and gain some stability, it might be possible for a taxonomy of criteria for evaluating teaching to be developed, comprising the core or most basic elements which are included in the various programs for evaluating teaching found throughout the state. With such a taxonomy available, it might be practical to devote considerable effort and expense to developing and validating instruments for measuring the factors which relate to the classified criteria. Thus, a common library of professionally developed and standardized instruments would be available for use by those schools which found them appropriate. In using these instruments it would still be possible for different schools to select different scales and assign different weights to those selected. Their use would contribute significantly to interpretability. In addition, the possibility for local schools to use only those scales which they deemed relevant and to apply unique weights to those selected provides for local initiative.

The most essential prerequisite for interpretability is sufficient knowledge on the part of the person who is analyzing the data of the conditions under which they were collected. Thus, a further step for enhancing interpretability would be to file information on these conditions with the data record. The conditions could include not only the time, place, and circumstances of the observation, but also relevant facts about the training of the observer.

The third condition necessary, if information on teaching is to be of value in implementing state educational leadership, is reliability. This term refers to the consistency between information on teaching which is collected and the behaviors or conditions which that information is assumed to represent. In other words, if precisely the same type of teacher behavior were to take place in two different schools and this behavior were recorded by observers in both schools, it should be possible to ascertain from the independently prepared observation records that the behavior in the two schools was identical. When the rating scales or other measures used provide a stable and veridical representation of the teaching to which they pertain, the resulting data are high in reliability.

Since reliability of measures is essential in local evaluation, whether or not the data are to be used elsewhere, the use of data collected in carefully developed local programs should ensure some measure of reliability. To increase this likelihood, local schools and school districts could be provided consultation and other assistance in developing scales and in training observers. It should be recognized, however, that such procedures will contribute to reliability only for measures taken within schools. They will not assure that measures taken with the same scale in two different school districts will be at all consistent. Further efforts would be needed such as programs which would bring together people from different schools for the purpose of discussion and practice in observing teaching processes and products.

Obviously, the problems of inter-school reliability and interpretability of data collected from various sources are closely intertwined. Therefore, the proposals made in the preceding discussion of interpretability are relevant here also. The development of a taxonomy of teaching criteria and of carefully constructed instruments for measuring the various elements of the criteria would certainly contribute to the reliability of data collected in more than one school or school system.

Much of the above discussion relates most directly to rating or other types of systematic observation of teacher or pupil performance. If tests are used to evaluate teaching in terms of product criteria, procedures for obtaining reliability and interpretability which have already been treated extensively in the literature on testing are immediately applicable.

Collecting and Storing the Data

To conduct the desired analyses, it is necessary to have information on teaching which is collected throughout the state assembled in such a manner that it will be accessible when it is needed. Because of the volume of the information and the complex problems inherent in selecting the information for any given analysis, the proposed system is designed with electronic data processing equipment in mind. The data which are collected would be stored on magnetic tape, magnetic discs, or in some other form allowing for access *via* computer.

The data collected from local school districts would include not only the information which represents measures of the teaching which has taken place, but also information which establishes its relevance and provides guidance in its interpretation. Since the system is intended to be used to provide information for use in making decisions relative to teacher education, information obtained from teacher education institutions pertaining to appropriate procedures and criteria for evaluating their graduates would also be obtained.

The data would be stored in a number of separate data files. Appropriate information could be selected from any one of these files or any combination of these files for analyses which might be conducted in reference to numerous specific questions. Descriptions of the proposed data files are given below.

Information on Criteria

The first two files would contain statements of criteria and procedures which are deemed appropriate for evaluating teaching in terms of the criteria. Information for these files would be obtained from all school districts in the state and from all programs for preparing teachers within the state.

Teacher education program evaluation criteria.--This file will be a composite record of the criteria which each of the teacher education institutions considers appropriate for evaluating the teaching of its graduates. A set of criteria could be applicable to all graduates of an institution, to graduates of a specific program within an institution, or only to individual teachers. The criteria would relate to the performance of teachers and/or to the outcomes of their teaching. Each set of criteria would be stated in a manner sufficiently specific for determining whether or not it would be appropriate to use them in conjunction with teaching performance or teaching product information obtained from the various local school districts where teachers to whom the criteria are applicable might be teaching.

School program evaluation criteria.--This file will contain a compilation of the criteria which local school systems consider relevant for evaluating teaching in their schools along with the procedures which are to be employed for evaluation. As with the criteria in the preceding file, a given set could apply to all teachers in an entire district, to teachers of a given subject, at a given level, in a specific school, teaching a specific type of class, or to an individual teacher only. The criteria would relate to the performance of teachers and/or to the outcomes of their teaching.

Status Information

The second group of files would contain information on the status of persons certified to teach in the state and on the schools in the state. This information would be compiled from certification and accreditation records. Information in these files would pertain to all teachers and to all schools.

School status information.--This file will include information on schools which is considered significant in evaluating the quality of the school program. It would relate to school organizations, enrollment, curriculum, and all other factors which are given official consideration when determining the eligibility of a school for accreditation.

Teacher status information.--This file will contain information on individual teachers which is considered relevant in predicting teaching competence. This would include those factors which are used to determine whether or not a baccalaureate graduate should be recommended for a teaching certificate, and whether or not such a certificate should be issued.

Evidence of Teaching

The third set of files would contain information on teaching performance and teaching products which have been observed and recorded. This information will be collected by local school personnel while carrying out local programs for evaluating teaching. It is not anticipated that all evaluation data collected by local schools would be forwarded to the state department of education for inclusion in these files. The state personnel would analyze information in the criteria files and then request from local school districts the information which would be usable for analysis in reference to specific questions of interest.

Teaching performance information.--This file will contain measures of the teaching performance of individual teachers. The measures on each teacher will be obtained with instruments and procedures which provide for interpretability, employing the criteria specified for that teacher in the file of school program evaluation criteria.

Pupil performance information.--This file will contain measures of the performance of pupils taught by teachers who are to be evaluated employing this type of criterion. The measures will, of course, be in a form which will allow them to be related to the applicable criteria.

Analyzing the Data

The general procedure proposed for analyzing the data involves two phases: the compilation of the file of data for a specific analysis and the analysis, itself. The general design for the analysis system is depicted in Figure 8.

The compilation of the analysis file will consist of selecting from the data files the information which will be analyzed. This will involve selecting records from individual files according to given selection criteria, and/or combining records or items from individual files to build new records.

The number of specific analyses which might be conducted employing the data in the files described is virtually unlimited. Some analyses which relate to the specific questions listed in Chapter I are considered below. A restatement of the applicable question or questions precedes each discussion.

(1) What criteria do individual colleges think are relevant for evaluating the teaching of graduates of their programs? Are these the same for all graduates? For all graduates in given teaching fields?

The information for answering this question would be contained in the file of teacher education program evaluation criteria. In order to

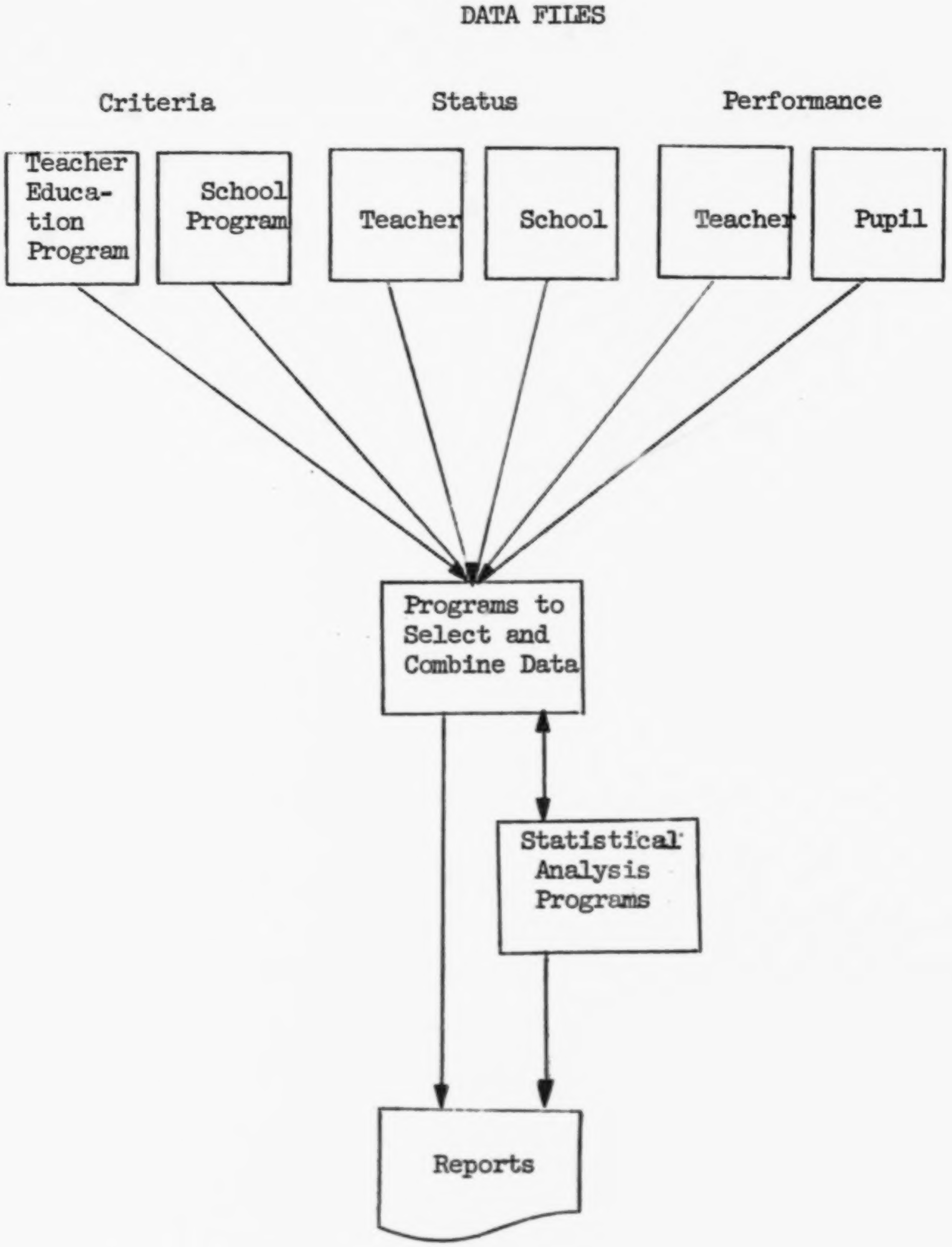

DATA FILES
Criteria
Status
Performance
Teacher Education Program
School Program
Teacher
School
Teacher
Pupil
Programs to Select and Combine Data
Statistical Analysis Programs
Reports

Fig. 8--General Design for the Data Analysis System.

answer the question posed, it would be necessary to develop a system for classifying the criteria, possibly according to the dimensions of teaching to which they relate. Examples of different dimensions stated in terms of teacher behavior might include those used by researchers: e.g., verbal influence of the teacher as analyzed by Flanders, logical use of languages as analyzed by Smith, or philosophic concepts related to teaching performance as described by Brown. Or, they might include more traditional categories such as maintaining discipline, using teaching aids, or providing for individual differences. Dimensions stated in terms of pupil performance might be described in terms of the categories in the Taxonomy of Educational Objectives: e.g., knowledge, comprehension, application, responding, valuing. As soon as the criteria are classified, the above question can be answered by tabulating the classification information. The answers can be provided in several forms: the information could be tabulated by subject specialty, by institution, by type of institution, by geographic area, or by whatever unit of analysis is called for by the specific problem at hand.

> (2) How does the teaching done by graduates of different types of teacher education programs measure up against the evaluation criteria considered appropriate by the teacher education institutions? By the local school districts?

The analyses relevant to this question would begin by identifying the teachers whose teaching is to be studied. This could be done with the file of teacher status information which would contain information on the programs of preparation completed by individual teachers. The criteria that have been established for evaluating these particular teachers would then be reviewed using both the teacher education program criteria file and the school program criteria file. Finally, usable information on those teachers from the file of teaching performance information and/or the file of pupil performance information would be analyzed.

> (3) What criteria do individual school districts or schools judge to be relevant for evaluating teaching in their classrooms? Are these the same for all or most teachers within a school district? Within a school? For all or most teachers of certain grades or subjects?

This question is almost the same as the first question, the difference being that it pertains to criteria established by local school districts while the earlier one referred to criteria established by teacher education institutions. The procedure would be basically the same as described earlier: (1) to identify the schools to be included in the population for analysis, (2) to classify the criteria, and (3) to tabulate the data so that they can be comprehended.

> (4) How does the teaching done by teachers in individual schools measure up against the evaluative criteria which are set forth as appropriate by the school district?

The first step in dealing with this question would be to determine which schools were to be selected and how the teachers from those schools were to be grouped. It would be necessary also to determine what phases of teaching were of interest. The necessary records from either the pupil performance or the teaching performance files, depending upon the type of criteria employed in the schools selected, could be used and the data could be tabulated or otherwise organized into a form which would provide for interpretability.

(5) How do the criteria specified for different segments of the teaching population (e.g., beginning teachers, music teachers, sixth grade teachers, etc.) compare?

This question would require that the segments of the teaching population to be analyzed first be identified, with the use of the teacher status file, and that the criteria established for evaluating the work of these people then be selected from the teacher education program criteria file and/or the school program criteria file. The analyses would consist of describing the data through tabulation or other techniques.

(6) How do evaluation criteria selected by school districts and evaluation criteria selected by colleges compare?

The first step in dealing with this question would be to select a population of teachers, possibly on the basis of their having completed selected teacher education programs. The teacher status file could be used for this purpose. Records for these teachers on the school program criteria file and the teacher education program criteria file could then be compared. The results of the comparison could be summarized.

(7) Is there more than one distinctive pattern of expectations for teachers in a given field?

The analyses for this question employ data from the two criteria files. The criteria which are applicable to teachers in the specific field under consideration would be selected and tabulated. This is the same type of analysis proposed in connection with the questions dealing with types of criteria specified by teacher education institutions or local school systems.

(8) What are the inservice education needs of various segments of the teaching population?

Information relevant to this question would be obtained by securing from the files of teacher education program evaluation criteria and school program evaluation criteria an indication of the dimensions of teaching which are considered relatively important for a group to which a program of inservice education would be directed. Such a group might be composed of teachers of a given subject or teachers in a given geographic area. Information would then be retrieved from the teacher performance file which would indicate the relative strengths and weaknesses of teachers in each of the significant aspects of teaching.

(9) What are the relationships between teacher status characteristics (viz., the things which are considered in recommending a baccalaureate candidate for a teaching certificate or in issuing a teaching certificate) and teaching (evaluated in terms of criteria established by local school districts)?

The first step in dealing with this question would be to determine which specific status factors and which specific dimensions of teaching are of interest. Then, records containing measures of the significant dimensions of teaching can be selected from the file of teaching performance data. The desired status data on the teachers represented can be selected from the file of teacher status information and the two sets of information can be compared.

(10) What relationships exist between school status characteristics (viz., the things which are given consideration in recommending a school for accreditation) and the teaching and learning which takes place in those schools.

This question would require the same basic procedures as the preceding one. It would first be necessary to identify the specific status factors and also the specific types of pupil or teacher performance which are of interest. After selecting the records containing the information to be analyzed, the data could be compared.

Initial Steps in Developing a State Program to Collect and Analyze Information on Teaching

Implementation of the system for collecting and analyzing information on teaching which is described in this chapter involves three phases: (1) developing in local schools or school districts carefully conceived programs for evaluating teaching, (2) developing communication and collaboration between schools and school districts which will make possible the combining of data from different evaluation programs, and (3) developing a central system for storing and analyzing the data collected. In the design of the total system, these phases are interrelated; therefore, it is proposed that they be pursued concurrently, rather than sequentially. Various ideas which warrant consideration in carrying out these phases are discussed below.

Inducing Local School Districts to Establish Evaluation Programs

Local programs for the evaluation of teaching are neither new nor uncommon. Surveys of teacher rating practices in public schools date back at least to 1915.[1] The most recent National Education

---

[1] Hazel Davis, "Evolution of Current Practices in Evaluating Teacher Competence," Contemporary Research on Teacher Effectiveness, ed. Bruce J. Biddle and William J. Ellena (New York: Holt, Rinehart and Winston, 1964), pp. 41-66.

Association survey of this type reports that a majority of large school systems and at least a fourth of the smaller school systems represented in its nationwide samples have formal programs for evaluating teaching.[1] Thus, the problem of getting local evaluation programs started is reduced considerably.

An obvious way to induce local school districts which do not have programs for the evaluation of teaching to establish them would be to pass a law or a regulation. However, such an action would not be likely to promote an attitude on the part of local school personnel which would result in the type of commitment necessary to develop high quality programs. Development of the desired evaluation programs would require conscientious local initiative in seeking a definition as to what, under specified circumstances, constitutes good teaching. Once this is defined, there is the equally challenging problem of developing objective procedures for collecting and analyzing the necessary evidence to determine whether good teaching has taken place. A tendency could develop to perform tasks of this type in a superficial manner if their sole purpose were to fulfill the requirements of a statute.

It might be advisable, however, to establish an official regulation or policy affirming the desirability of locally developed evaluation programs. Efforts could also be made to support pilot programs which could demonstrate the application of various approaches.

---

[1] Hazel Davis, Evaluation of Classroom Teachers, Research Report 1964-R14 (Washington, D. C.: Research Division, National Education Association, 1964), p. 14.

Eventually, when most school districts have developed sound programs, it might be advisable to pass a law or regulation to pull along those who had not done so, if it were felt that all schools should have a formal program for evaluating teaching.

In cases where statutes are in effect which specify the forms and procedures to be used for evaluating teaching, the evaluation laws would, of course, have to be repealed or amended to allow local school districts to develop their own evaluation programs. While conditions vary from state to state, an amendment would probably be preferable to repealing the law completely. This is suggested to avoid conveying the impression that the state no longer thinks it important to evaluate teaching. The amended law would need to provide only (1) that all teachers be evaluated annually, at least, employing criteria developed by the professional personnel within the district, and (2) that, upon request, criteria and procedures employed in each district and data collected in the evaluation program be submitted to the state department of education. The first stipulation would offer assurance that the state is still interested in protecting its citizenry from incompetent teaching. The second provision would officially give the state access to the data needed for the information system, but would not require schools to furnish the state department of education with information which it was not prepared to put to use.

Pilot Projects to Develop Exemplary Programs for Evaluating Teaching

The conceptualization of evaluation criteria and the translation of those criteria into operational terms as required by the proposed evaluation system are demanding tasks. To increase the likelihood of developing programs of exceptional quality, it is proposed that a number of pilot projects be instituted which would make use of resources beyond those that can normally be committed in a single school or district. For such projects, supplemental financing could be made available to allow certain school district personnel to devote a greater proportion of their time to the evaluation project and also to bring in consultative assistance from outside the district. Consultants might include scholars in educational philosophy, authorities on the principles and techniques of teaching, and experts in research and measurement. Pilot projects could be initiated by colleges and universities, by school districts, or jointly by colleges or universities and school districts.

Each of the projects would identify the dimensions of teaching to be incorporated in its criteria, define the criteria, develop and validate the necessary measurement instruments, and demonstrate the application of the evaluation program in an ongoing school setting. Instruments could be demonstrated and tested in the pilot school and in other schools which volunteered to participate.

The ultimate aim of the pilot projects would not be to impose upon all schools certain criteria or procedures for evaluating teaching. Nor would the projects be intended in any manner to restrict other schools in the development of their own criteria and measuring devices. The goal would be simply to elevate the standards of evaluation in given schools and to make available information regarding procedures employed.

Disseminating Information on Newly Developed Programs for the Evaluation of Teaching

Although it is not proposed that methods for measuring and evaluating teaching developed in pilot projects be imposed upon any school, it is assumed that if the methods have merit they will be adopted. This is, of course, contingent upon the prospective adoptors being introduced to the newly developed evaluation systems. A number of procedures for communicating this information might usefully be employed. These include publications; conferences or workshops at the state, regional, county and local school levels; and informal consultation.

The publications could include magazine articles or brochures describing the various approaches to the evaluation of teaching being taken in different pilot projects and their stages of development. When an evaluation system has been developed and demonstrated, detailed bulletins could be published describing the rationale employed in the development of the evaluation system and the specific procedures for applying it. Data could also be given regarding the reliability and validity of the measuring techniques employed.

Conferences or workshops of a variety of types could be scheduled for a number of different purposes. Conferences could be scheduled in advance of the pilot projects to arouse interest in participation. As the pilot projects progress, conferences could be held to demonstrate, compare, and evaluate the approaches which are being taken in the various projects. This may also point out the need for additional projects. When an evaluation program reaches a stage calling for testing of the procedures at other schools, conferences could be held to demonstrate the applicability of those procedures. When a project has validated its measuring procedures, conferences could be held to demonstrate them to persons who might be interested in using them. When a school has chosen to adopt a particular evaluation system, a workshop can be scheduled to instruct teachers and administrators in its use.

Virtually all of the activities described in the preceding paragraphs could also be accomplished through informal meetings and consultation. The particular approach selected would depend upon the specific circumstances in each situation.

Developing the Central Data System

The planning of the central data system should begin as soon as the decision is made to develop a state system for organizing and analyzing information on teaching. There should be continual communication between the persons designing the data system and those in local school systems to make possible optimum adaptability of the data system to the anticipated data and, when feasible, to provide for optimum adaptability of the anticipated data to the data system. This is not to say that evaluation programs should be designed for the convenience of the data processors; if, however, certain modifications of evaluation procedures which would increase data processing capabilities could be accomplished without affecting the underlying rationale or the overall quality of the evaluation program, they should certainly be considered. It is particularly important that communication be maintained between schools carrying out pilot projects and the designers of the central information system since it is assumed that procedures from the successful projects will be adopted in numerous additional schools.

The actual building of the files must wait until data are available, although fictitious data could be used to try out processing programs and procedures. Data for the status files would be available immediately, however, since this information is normally contained in state department of education records. Information for criteria files would be available before information for performance files, as the data for the latter are dictated by the criteria.

CHAPTER V

RECAPITULATION AND REFLECTIONS

This study was instituted to delineate the conditions under which statewide programs for measuring and evaluating teaching can yield optimum information for use by the states in carrying out their educational leadership responsibilities. The responsibilities under consideration relate particularly to three state services administered through state departments of education: the education of teachers, the certification of teachers, and the accreditation of schools.

Recapitulation

Chapter I discusses the types of activities carried out by state governments in their efforts to provide for high quality teaching in elementary and secondary schools. These include (1) developing and administering procedures to insure the quality of programs for the preparation of teachers, (2) developing and administering standards which will allow only qualified people to hold teaching positions, and (3) developing and administering programs for promoting teacher growth and eliminating substandard teaching situations. State educational leadership consists of devising ways to improve these activities. Comprehensive information on the teaching which is taking place in the schools would be of great value in implementing this leadership.

It is noted that virtually no systematic use of such information is presently made by state governments. This could be attributable to the fact that much information on teaching which has been or might be collected is ambiguous. There is no widespread agreement as to what criteria sould be used for evaluating teaching. Therefore, confusion may result when two people analyze the same evidence because they are using different criteria. Much research has been conducted to determine the factors which can account for effectiveness of ineffectiveness in teaching. This has generally not been fruitful. A further complicating factor relates to the tenuous relationship between teaching and the subsequent behavior of pupils. The influence of the teacher is only one of the many effects which contribute to pupil learning.

In Chapter II, the three statewide programs for evaluating teaching which are presently operative are described. The State of Pennsylvania has prescribed a form and procedures for rating teachers. Satisfactory ratings on the form are required in order for a teacher to be eligible for tenure and in order for him to have his certificate renewed. An unsatisfactory rating, along with supporting evidence, provides a legal basis for dismissing a tenure teacher on grounds of incompetence. The Pennsylvania program is not designed to provide information which can be used in making decisions relative to teacher education, certification and accreditation.

The statewide evaluation program in Hawaii has as its stated purpose the improvement of teaching in the schools of that State. It is a program similar to one which might be developed in a large school district. The state prescribes a form and procedures; principals observe the teachers and, following the visits, review the ratings in a conference with the teachers. The program for evaluating the work of tenure teachers performs no official regulatory function. However, as in Pennsylvania, satisfactory ratings are required for a probationary teacher to become eligible for tenure. The Hawaii program is not designed to provide information which can be used to review the effectiveness of state educational policies and programs.

In Florida, all teachers are evaluated annually using a form prescribed by the State. The results are placed on file in the State Department of Education in a manner which allows them to be summarized and analyzed using electronic data processing equipment. While no official purposes have been specified for this program, it is organized in a manner which should provide the kind of information which could be used to make decisions regarding state policies for improving teaching. However, an examination of data collected indicates various shortcomings which restrict considerably interpretations that can be made using this information.

Chapter III discusses approaches to the evaluation of teaching. The process of evaluation, as described in the present study, consists of comparing evidence with criteria. Consequently, the nature of the evidence which is relevant is determined by the criteria. Three types of criteria--product, process, and presage--are described. Either the product type or the process type can be used in evaluating; presage criteria are, in reality, predictors. The selection of criteria constitutes a value judgment which may be quite arbitrary. However, a person who is familiar with the situation to which the criteria will apply is normally in a better position to make a judgment as to what constitute acceptable criteria.

The process of collecting evidence and comparing it with criteria is viewed along three dimensions. They relate to (1) the nature of the situation in which the evidence is collected (varying from "normal" to "constructed"), (2) the agent responsible for collecting the evidence (with the performer, himself, represented at one extreme and observer at the other), and (3) the point at which the evidence is compared with the criteria (varying from "while observing" to "independently").

The applicability of various techniques for collecting evidence and comparing it with criteria are discussed. These include self reports and self evaluation, rating and rating scales, systematic observation with evaluation performed independently, and testing. There are conditions under which each of these techniques might be applicable in a statewide program for obtaining information to implement educational leadership.

Chapter IV describes an approach for organizing and analyzing information on teaching to implement state educational leadership. The effectiveness of such a program depends first upon three conditions which must be met by the information used; they are, relevance, interpretability, and reliability. Relevance pertains to the relationship between the information collected and the circumstances which state educational leadership is intended to effect. Because this leadership is aimed primarily at helping local school districts to carry out programs which meet the needs of their citizens, it is proposed that the most relevant information can be obtained from carefully designed local evaluation programs. This, however, could lead to problems of interpretability.

Interpretability is the second essential condition for an effective program. Information which is interpretable can be summarized and analyzed in ways which will yield information that can be used for the desired purposes, in this case, for the implementation of state educational leadership. Attempting to combine data collected in different places under different circumstances is certain to bring about problems of interpretability. To deal with these, it is proposed that communication be developed between schools using similar criteria, that pilot programs for the development of exemplary evaluation systems be established and the results widely disseminated, and that a taxonomy of criteria including the core or most basic elements which are included in evaluation programs throughout the state be compiled. The taxonomy could be the starting point for the development and standardization of instruments for measuring various elements of teaching.

Reliability, the third essential condition for an effective program refers to the consistency between the information on teaching which is collected and traits, behaviors, or conditions which that information is assumed to represent. Reliability results when carefully developed instruments and procedures are used.

Six different files are proposed for collecting and storing the information in the state department of education. They would contain information on (1) teacher education program criteria, (2) school program criteria, (3) teacher status information, (4) school status information, (5) teaching performance information, and (6) pupil performance information. These files would be maintained in a form which would allow access via computer.

The general procedure for analyzing the data involves two phases. The first consists of selecting from the files the data for analysis. The data could come from any one file or any combination of files. The selected data would be incorporated into a new file. The second phase consists of the analysis, itself. This would utilize descriptive or inferential statistical techniques. Brief illustrations of applications for the analysis system are presented in Chapter IV.

Reflections

If the specifics of the proposals made in this study are pushed into the background, the general message can be summarized in the form of two proposals. First, state governments should broaden their commitment to institutional research as a basis for decision-making in the area of education. And second, educators in the elementary and secondary schools should begin to codify their professional wisdom.

Codifying the Wisdom of Educational Practitioners

The proposal for codifying the knowledge of practitioners developed in the present study as an outgrowth of the institutional research idea. It resulted from the concept that evaluation must be based upon criteria and the stipulation that neither the teaching processes to be employed nor the teaching products to be sought should be dictated by the state to local school personnel. This means that, if evaluation is to take place, local school personnel must establish the criteria. The establishment of criteria by such personnel would constitute a codification of their professional wisdom.

The resulting statements would set forth the factors which, in the judgment of practicing educators, determine sound educational practice. It would not be a superficial document as might result from a simple survey or Q-sort. Since the statement of criteria would form the official basis for evaluating the work of teachers, it would be reasoned out in a manner which would strike for the essence of the issues. As the criteria were applied, the codification would be continually re-evaluated, revised, and improved. As communication and collaboration between schools increased, the body of knowledge developed by professional practitioners would become more refined.

Lortie[1] declares that a codified body of knowledge representing the best thinking of skilled professional practioners is something which is lacking in the field of education. This is not the case, however, in the professions of medicine, law, and architecture. He notes that, while the professional subject matter for the teacher contains contributions from philosophers and psychologists, it does not include a body of knowledge codified by educational practitioners. In medicine, law, and architecture there are such courses, based upon the cumulated wisdom of practitioners.[1]

---

[1] Dan C. Lortie, "Teacher Socialization: The Robinson Crusoe Model," The Real World of the Beginning Teacher (Washington, D. C.: National Commission on Teacher Education and Professional Standards, 1966), pp. 54-66.

Lortie also observed that there exists a chasm between schools of education and teachers in the elementary and secondary schools. He asserts that teachers do not attribute the same importance to professional schooling as do physicians, lawyers, and architects.[2] Research results documenting the gap between education professors and teachers are reported by Joyce,[3] who found that attitudes toward teaching held by beginning teachers are more in agreement with those held by the general public than with those held by professors of education.

The codifying by practitioners of a body of practical knowledge could contribute to the improvement of teacher education and also to the narrowing of the breach between practitioners and professors.

---

[1] Lortie gives Internal Medicine, Torts, and Principles of Design as examples of this type of course. Ibid., p. 60.

[2] Ibid., p. 62.

[3] Bruce R. Joyce, "The Social Climate of Teacher Education," Journal of Teacher Education, Vol. 14 (June, 1963), pp. 179-183.

This knowledge would give beginning teachers invaluable support as they commence their professional duties. Excerpts from this body of knowledge would certainly be adopted by schools or departments of education and integrated into their programs. Hopefully, the ultimate result would be the wedding of the best elements from both the professors and the practitioners. One effect of this would be a more prominent place for elementary and secondary classrooms in programs of teacher education.

It is often stated that professional education for teachers should become more closely associated with the classroom. Typical recommendations are that longer student teaching periods be established, that internship programs be developed, that students be assigned case studies or other projects which will bring them into the schools, or that instructors in graduate courses slant their presentations more directly toward the problems of the teachers enrolled.

It is seldom suggested that the wisest and most skillful teachers perform the difficult task of codifying their professional wisdom so that it can benefit both them and their associates. Yet, it seems unnecessary that each first-year teacher should have to begin anew to develop such a body of professional wisdom. It also seems unreasonable that a professor of education should have to relate his psychology or philosophy to chance examples when a body of professional information codified by educational practitioners might be available to form a basis for discussion. The codification of the knowledge of practicing educators could serve to bring teacher preparation and teaching closer together. It would provide for an easier induction of the beginning teacher and a basis for communication between teachers and education professors from which they could proceed to work together to improve teaching.

Broadening the State Commitment to Institutional Research

The present study began by articulating a case for institutional research, without using the term. Institutional research is intended to provide data which makes it possible for administrators or leaders to make intelligent decisions. It represents a rational, rather than arbitrary, approach to decision-making. It is widely practiced in certain phases of state school administration. On matters relating to educational finance, such as cost of certain programs or tax-paying ability in certain areas or of certain segments of the population, institutional research is the rule, rather than the exception. Institutional research is also practiced extensively in anticipating needs for new facilities. However, as was pointed out in Chapter I, very little use is made of institutional research in developing policies for teacher education, certification, and accreditation.

The primary proposal in this study is that attempts be made to take a rational approach to these matters, as is typically taken in the domains of finance and facilities. Optimum decisions are more likely to result when as much relevant information as possible can be brought to bear.

Some Side Effects

There are at least two areas not yet mentioned in which implementation of the above general proposals would contribute to the improvement of teaching. These areas represent functions which are, in fact, more fundamental in improving teaching than are the state services toward which the information system is directed. The first relates to research on teaching; the second to stimulation of local leadership for the improvement of instruction.

In the area of research, the large volume of carefully collected measures and evaluations of teaching would make data available for numerous correlation studies or other investigations employing pre-experimental or quasi-experimental designs.[1] These would include studies probing relationships between various status characteristics of teachers and various types of performance of teachers or between various types of performance of teachers and various teaching products.

It is also expected that implementation of the system would lead to improved understanding and respect by school personnel of the researcher's approach. This would result from continual efforts by teachers and administrators to define their criteria and to employ measures for determining the extent to which they have been met. This is the same basic problem which the research worker encounters repeatedly throughout his career.

---

[1] Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally and Company, 1963), pp. 171-246.

The understanding which is expected to result should make the schools more accessible for use as laboratories for experimental studies. Moreover, it seems likely that the greater sophistication of the school people would contribute to better experimental controls and better research results.

Instructional leadership at the local level would also be affected by the greater understanding of research. The search for criteria and for methods of applying them could make school personnel more alert and receptive to research results which might provide assistance or guidance.

However, the greatest stimulus to instructional improvement within the local schools would almost certainly result from the initiative taken by local school personnel to define and redefine the criteria which they wish to employ in evaluating teaching. With definite criteria at hand and definite evidence as to the extent to which the criteria are being met, more decisive measures can be taken to improve teaching.

Conclusion

This study has proposed the development of an information system which is capable of providing facts which can be used by the state in strengthening educational leadership. The system is designed to allow the state to make decisions relative to policies for teacher education, teacher certification, and school accreditation on a more rational basis. The implementation of the system, however, would provide some additional benefits which are, in reality, more significant than state regulations and policies for improving teaching. These include (1) increasing the potential contributions of practitioners to teacher education, (2) stimulating educational research and the application of research findings, and (3) stimulating local leadership for the improvement of instruction.

APPENDICES

DIVISION OF TEACHER EDUCATION, CERTIFICATION AND ACCREDITATION
FLORIDA STATE DEPARTMENT OF EDUCATION
TALLAHASSEE 32304

Statewide Teacher Evaluation Questionnaire

Has your state within the last ten years conducted a statewide teacher evaluation project? yes___ no___

Do you conduct regular annual statewide teacher evaluations? yes___ no___

Indicate the years of your three most recent statewide teacher evaluation projects.

1966 .1965 1964 1963 1962 1961 1960 1959 1958 1957

Answer the following questions for each of your three most recent statewide teacher evaluation projects.

	most recent project	earlier project	earliest of the three projects
Was it conducted in an effort to recognize and promote meritorious teaching?	yes___ no___	yes___ no___	yes___ no___
Was it conducted in an effort to recognize and eliminate incompetent teaching?	yes___ no___	yes___ no___	yes___ no___
Was it conducted for the purpose of accumulating data for research?	yes___ no___	yes___ no___	yes___ no___
Was it conducted for purposes other than those mentioned above?	yes___ no___	yes___ no___	yes___ no___

PLEASE ENCLOSE POLICY STATEMENTS, EVALUATION INSTRUMENTS, PROJECT REPORTS, OR ANY OTHER AVAILABLE MATERIAL WHICH IS RELEVANT TO YOUR STATEWIDE TEACHER EVALUATION PROJECTS.

____
(person completing questionnaire)

____
(title)

FLOYD T. CHRISTIAN
SUPERINTENDENT

STATE OF FLORIDA
DEPARTMENT OF EDUCATION
TALLAHASSEE 32304

August 1, 1966

We would like to request your assistance in obtaining information on statewide teacher evaluation practices in your state. This information is solicited as part of a larger study of practices and potential for statewide teacher evaluation programs.

Statewide teacher evaluation is defined here as an attempt by the state government (or its agent) to make a systematic composite assessment of the quality of teaching or teachers within the entire state.

Your cooperation is greatly appreciated.

Sincerely yours,

Wm. Cecil Golden, Director
Division of Teacher Education,
Certification and Accreditation

WCG:ad

Enclosures: Questionnaire
Return Envelope

FLOYD T. CHRISTIAN
SUPERINTENDENT

STATE OF FLORIDA
DEPARTMENT OF EDUCATION
TALLAHASSEE 32304

August 29, 1966

About four weeks ago we requested information from you relative to statewide teacher evaluation practices. Since I have not yet received your reply, I am enclosing a second copy of the questionnaire in the event that the first might have been mislaid.

I hope that you will return the questionnaire. Your assistance is greatly appreciated.

Sincerely,

Wm. Cecil Golden, Director
Division of Teacher Education,
Certification and Accreditation

WCG:ad

Enclosures: Questionnaire
Return Envelope

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

Armstrong, W. Earl, and Stinnett, T. M. A Manual on Certification Requirements for School Personnel in the United States. Washington, D. C.: National Commission on Teacher Education and Professional Standards, National Education Association, 1964.

Amidon, Edmund, and Flanders, Ned A. The Role of the Teachers in the Classroom: A Manual for Understanding and Improving Teachers' Classroom Behavior. Minneapolis: Paul S. Amidon and Associates, 1963.

Bloom, Benjamin S. (ed.). Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay Company, Inc., 1956.

Davis, Hazel. Evaluation of Classroom Teachers. Research Report 1964-R14. Washington, D. C.: Research Division, National Education Association, 1964.

Ellena, William J., Stevenson, Margaret, and Webb, Harold V. (eds.). Who's a Good Teacher? Washington, D. C.: American Association of School Administrators, 1961.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1954.

Harris, Chester W. (ed.). Encyclopedia of Educational Research. New York: The Macmillan Company, 1960.

Krathwohl, David R., et al. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, Inc., 1964.

Mayor, John R., and Swartz, Willis G. Accreditation in Teacher Education: Its Influence on Higher Education. Washington, D. C.: National Commission on Accrediting, 1965.

Ryans, David G. Characteristics of Teachers: Their Description, Comparison and Appraisal. Washington, D. C.: American Council on Education, 1960.

Thurston, Lee M., and Roe, William H. State School Administration. New York: Harper and Brothers, 1957.

Travers, Robert M. W. An Introduction to Educational Research. 2nd edition. New York: The Macmillan Company, 1965.

Articles and Periodicals

Akerman, Walter I. "Teacher Competence and Pupil Change," Harvard Educational Review, Vol. 24 (Fall, 1954), pp. 273-289.

Anderson, Harold M. "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, Vol. 23 (September, 1954), pp. 41-71.

Barr, Arvil S. "The Measurement and Prediction of Teaching Efficiency; A Summary of Investigations," Journal of Experimental Education, Vol. 16 (June, 1948), pp. 203-283.

Barr, Arvil S., Eustice, David E., and Noe, Edward J. "The Measurement and Prediction of Teacher Efficiency," Review of Educational Research, Vol. 25 (June, 1955), pp. 261-269.

Barr, Arvil S., and Jones, Robert E. "The Measurement and Prediction of Teacher Efficiency," Review of Educational Research, Vol. 28 (June, 1958), pp. 256-264.

Barr, Arvil S., et al. "The Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations," Journal of Experimental Education, Vol. 30 (September, 1961), pp. 5-156.

Biddle, Bruce J. "The Integration of Teacher Effectiveness Research," Contemporary Research on Teacher Effectiveness. Edited by Bruce J. Biddle and William J. Ellena. New York: Holt, Rinehart and Winston, 1964, pp. 1-40.

Bradley, Ruth, et al. "A Design for Teacher Evaluation," The National Elementary Principal, Vol. 43 (November, 1963), pp. 32-27.

Brown, Bob Burton. "Bringing Philosophy into the Study of Teacher Effectiveness," Journal of Teacher Education, Vol. 17 (Spring, 1966), pp. 35-40.

Classroom Interaction Newsletter. Vol. 1 (May, 1966), pp. 1-2.

Cogan, Morris L. "Research on the Behavior of Teachers: A New Phase," Journal of Teacher Education, Vol. 14 (September, 1963), pp. 238-243.

Combs, Arthur W. "Can We Measure Good Teaching Objectively?" *NEA Journal*, Vol. 53 (January, 1964), p. 34.

Davis, Hazel. "Evolution of Current Practices in Evaluating Teacher Competence," *Contemporary Research on Teacher Effectiveness*. Edited by Bruce J. Biddle and William J. Ellena. New York: Holt, Rinehart and Winston, 1964, pp. 41-46.

Domas, S. J., and Tiedeman, D. V. "Teacher Competence: An Annotated Bibliography," *Journal of Experimental Education*, Vol. 19 (December, 1950), pp. 101-218.

Fattu, Nicholas A. "What Research Says About Teacher Effectiveness," *NEA Journal*, Vol. 50 (October, 1961), pp. 55-56.

———. "Research on Teacher Evaluation," *The National Elementary Principal*, Vol. 42 (November, 1963), pp. 119-125.

Gordon, Ira J "Determining Teacher Effectiveness," *Educational Leadership*, Vol. 20 (November, 1963), pp. 119-125.

Griffiths, Daniel E. "Administrative Theory and Change in Organizations," *Innovation in Education*. Edited by Matthew B. Miles. New York: Bureau of Publication, Teachers College, Columbia, University, 1964, pp. 425-436.

Hemphill, John K. "Administration as Problem Solving," *Administrative Theory in Education*. Edited by Andrew W. Halpin. Chicago: Midwest Administration Center, University of Chicago, 1958, pp. 96-106.

Howsam, Robert B. "Teacher Evaluation: Facts and Folklore," *The National Elementary Principal*, Vol. 42 (November, 1963), pp. 6-18.

Joyce, Bruce R. "The Social Climate of Teacher Education," *Journal of Teacher Education*, Vol. 14 (June, 1963), pp. 179-183.

Kelley, J. T. "Teacher Evaluation," *Florida School Bulletin*, Vol. 24 (December, 1961), pp. 23-28.

Lipham, James M. "Leadership and Administration," *Behavioral Science and Educational Administration*. The Sixty-third Yearbook of the National Society for the Study of Education, Part II. Edited by Daniel E. Griffiths. Chicago: The University of Chicago Press, 1964, pp. 119-141.

Lortie, Dan C. "Teacher Socialization: The Robinson Crusoe Model," *The Real World of the Beginning Teacher*. Washington, D. C.: National Commission on Teacher Education and Professional Standards, 1966, pp. 54-66.

Medley, Donald M., and Mitzel, Harold E. "A Technique for Measuring Classroom Behavior," *Journal of Educational Psychology*, Vol. 49 (April, 1958), pp. 86-92.

_______. "Measuring Classroom Behavior by Systematic Observation," *Handbook of Research on Teaching*. Edited by N. L. Gage. Chicago: Rand McNally and Company, 1963, pp. 247-238.

_______. "The Scientific Study of Teacher Behavior," *Theory and Research in Teaching*. Edited by Arno A. Bellack. New York: Bureau of Publications, Teachers College, Columbia University, 1963, pp. 79-90.

Mitzel, Harold E. "Can We Measure Good Teaching Objectively?" *NEA Journal*, Vol. 53 (January, 1964), pp. 35-36.

Remmers, H. H. "Rating Methods in Research on Teaching," *Handbook of Research on Teaching*. Edited by N. L. Gage. Chicago: Rand McNally and Company, 1963, pp. 329-378.

Remmers, H. H., *et al.* "Report of the Committee on the Criteria of Teacher Effectiveness," *Review of Educational Research*, Vol. 22 (June, 1952), pp. 238-263.

Remmers, H. H., *et al.* "Second Report of the Committee on Criteria of Teacher Effectiveness," *Journal of Educational Research*, Vol. 46 (May, 1953), pp. 641-658.

Rabinowitz, William and Travers, Robert M. W. "Problems of Defining and Assessing Teacher Effectiveness," *Educational Theory*, Vol. 3 (July, 1953), pp. 212-219.

Rose, Gale W. "Performance Evaluation and Growth in Teaching," *Phi Delta Kappan*, Vol. 45 (October, 1963), pp. 48-53.

Ryans, David G. "Assessment of Teacher Behavior and Instruction," *Review of Educational Research*, Vol. 33 (October, 1963), pp. 415-441.

_______. "Theory Development and the Study of Teacher Behavior," *Journal of Educational Psychology*, Vol. 47 (December, 1956), pp. 462-475.

Soar, Robert S. "Observation Systems and Research on the Teaching-Learning Process," *Classroom Interaction Newsletter*, Vol. 2 (November, 1966), pp. 1-3.

Turner, Richard L., and Fattu, Nicholas A. "Skill in Teaching, a Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research," Bulletin of the School of Education, Indiana University, Vol. 36 (May, 1960), pp. ii-vii, 1-40.

Yamamoto, Kaoru. "Evaluating Teacher Effectiveness: A Review of Research," Journal of School Psychology, Vol. 2 (January, 1963), pp. 60-71.

Miscellaneous Reports

[Brown, Bob Burton]. Teacher Competence Project. Madison, Wisconsin: School of Education, The University of Wisconsin, 1964. (Mimeographed.)

Hite, F. Herbert. "Effects of Reduced Loads and Intensive Inservice Training upon the Classroom Behavior of Beginning Elementary Teachers." Cooperative Research Project No. 2973. Olympia, Washington: Washington Superintendent of Public Instruction, 1966.

Kersh, Bert Y. "Classroom Simulation: A New Dimension in Teacher Education," Title VII Project Number 886, National Defense Education Act of 1958. Monmouth, Oregon: Teaching Research, Oregon System of Higher Education, 1963.

Morsh, J. E., and Wilder, E. W. "Identifying the Effective Instructor: A Review of Quantitative Studies, 1900-1952," Research Bulletin No. AFPTRC-TR-55-44. San Antonio, Texas: United States Air Force Personnel and Training Center, 1954.

"Personnel Memo No. 318 (1965-66)." Honolulu, Hawaii: State of Hawaii, Department of Education, October 22, 1965. (Mimeographed.)

Rasley, Charles W. "TER, A Pillar of Strength: A Current View of Teacher Effectiveness Rating as Used in the Public Schools." A Survey Research Study, Division of Education in the Graduate School, Southern Illinois University, Edwardsville, 1966.

Smith, B. Othanel, and Meux, Milton. "A Study of the Logic of Teaching." Cooperative Research Project Number 258, Trial Edition, Urbana, Illinois: Bureau of Educational Research, College of Education, University of Illinois, 1962.

Unpublished Materials

Brown, Bob Burton. "Teacher's Classroom Behavior." Gainesville, Florida: Teacher Competence Research Project. College of Education, University of Florida, undated.

Carter, Fletcher Fairwick. "Selected Aspects of Pre-Service Preparation and Prior Experience of High and Low Merit-Rated Junior College Teachers in Florida." Ph.D. dissertation, Florida State University, Tallahassee, 1964.

Cooper, James M. "A Performance Curriculum for Teacher Education." The Second Annual Florence B. Stratemeyer Lecture read before the Annual Meeting of the Association for Student Teaching, Chicago, February 16, 1967.

Daniel, K. Fred. "A Catalog of Analysis of Variance Pilot Studies." Research Report, Multi-State Teacher Education Project, State Department of Education, Tallahassee, Florida, 1966.

________. "A Study of the Objectivity of the Florida Teacher Evaluation Form." Research Report, Multi-State Teacher Education Project, State Department of Education, Tallahassee, Florida, 1967.

________. "Effective Teaching and Educational Values." Paper read before the Eighteenth Annual Meeting, Philosophy of Education Society, Southeastern Region, Tuscaloosa, Alabama, February 4, 1966.

Garrison, Harry L. "Evaluation of Teaching and Learning." Ed.D. dissertation, Standord University, Stanford, California, 1964.

Gerlock, Donald E. "An Analysis of Administrators' Evaluations of Selected Professionally and Provisionally Certificated Secondary School Teachers." Ed.D. dissertation, Florida State University, Tallahassee, 1964.

McTeer, John Hugh. "A Study of the Relationship of Teacher-Principal Likenesses and the Principals' Ratings of Teacher Effectiveness." Ed.D. dissertation, Florida State University, Tallahassee, 1963.

Packer, Morton Alfred. "A Study of the Relationship Between Teachers' Self-others Acceptance and the Principals' Ratings of These Teachers." Ed.D. dissertation, Florida State University, Tallahassee, 1964.

Reagan, Gerald M. "Toward a More Justifiable Theory for the Evaluation of Teachers and Teaching," Ph.D. dissertation, Michigan State University, East Lansing, 1964.

Wurzbach, Edward G. "A Three-County Study of Teachers' State Evaluation Scores in Relation to their Experience and Type of Degree." M. S. thesis, Florida State University, Tallahassee, 1962.

Yildirim, Cemal. "An Analytic Model for Evaluation of Teacher Competence." Ph.D. dissertation, Indiana University, Bloomington, 1963.

VITA

VITA

K. Fred Daniel, the author of this study, was born on February 7, 1936. His higher education has been at East Texas State College, where he earned bachelors and masters degrees, and at Florida State University where he earned the Ph.D.

He is presently employed in the Florida State Department of Education as the state director for the Multi-State Teacher Education Project. His previous professional positions were (1) coordinator of the student teaching program at Southwest Missouri State College (1963-65), (2) supervisor of off-campus student teaching at Bemidji State College (1962-63), (3) high school principal in Claremont, Minnesota (1961-62), and (4) music instructor at University High School, University of Minnesota (1958-61).

The Multi-State Teacher Education Project

The Multi-State Teacher Education Project (M-STEP) was initiated to strengthen the contributions of state departments of education to teacher education. This is being done through innovative pilot projects being carried out in each of the seven M-STEP states. The M-STEP design was adopted to enhance cooperation between the participating states in planning, analyzing, and evaluating the pilot projects and also in disseminating information on activities and results. In Florida, pilot projects intended to help teacher education institutions and elementary and secondary schools to obtain information which they can use in improving preservice and inservice teacher education are now underway. These projects have taken two forms: (1) bringing people together in the State to exchange ideas, and (2) developing techniques for analyzing data collected by the State Department of Education to provide usable information.

Project Coordinating Board

W. T. Boston, Assistant Superintendent, Certification and Accreditation, Maryland State Department of Education

Eugene Richardson, Consultant, Higher Education and Certification, Michigan State Department of Education

George W. Hopkins, Director, Division of Teacher Education and Certification, South Carolina State Department of Education

N. Blaine Winters, Administrator, Division of Teacher Personnel, Utah State Board of Education

Wendell C. Allen, Assistant Superintendent for Teacher Education and Certification, Washington State Department of Education

Genevieve Starcher, Director, Division of Teacher Preparation and Professional Standards, West Virginia State Department of Education

W. Cecil Golden, Assistant Superintendent for Teacher Education, Certification and Accreditation, Florida State Department of Education

State Project Coordinators

Jerry E. Chapman, Michigan
William H. Drummond, Washington
John B. Himelrick, West Virginia
Boyd Israel, South Carolina
Arthur T. Kulick, Maryland
Vere A. McHenry, Utah
K. Fred Daniel, Florida

Central Office Staff

Howard E. Bosley, Director
Charles K. Franzen, Associate Director

The Project Central Office is located in Suite 205, 1101 St. Paul Street,
Baltimore, Maryland 21202